Homework on Shannon-type inequalities

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 $S_1 \cap S_2^c \triangleq \phi$, where $\phi$ 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 $A \cap B = \phi$. 


Solution: The sets are \( \{ S_1 \cap S_2, S_1 \cap S_2^c, S_1^c \cap S_2, S_1^c \cap S_2^c \} \).

(b) 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

\[
\mathcal{U} \triangleq \{ H(X_1|X_2), I(X_1;X_2), H(X_2|X_1) \}
\]  

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.

Solution: By the properties of entropies and mutual information we can show that every information measure on \( X_1 \) and \( X_2 \) (or one of them) can be written as a sum of elements in \( \mathcal{U} \), with coefficients that are 0 or 1. Hence, it is a linear combination with non-negative coefficients.

(c) 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, \( \mu^*(S_1 \cup S_2) \to H(X_1, X_2) \).

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 \( \mu^* \).

Solution:
\( \cup \to \), and \( \mu^* \to H \)
\( \cap \to \); and \( \mu^* \to I \)
and $\mu^* \rightarrow H/I$.

If an intersection is involved, then $\mu^* \rightarrow I$.

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 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

$$h = [H(X_1), H(X_2), H(X_1, X_2), H(X_3), H(X_1, X_3), H(X_2, X_3), H(X_1, X_2, X_3)]^T \quad (2)$$

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

$$H(X_\alpha | X_\beta) \quad (3)$$

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

**Solution:** Conditional entropy can be written as linear combination of unconditional entropy, using the following identity:

$$H(X_\alpha | X_\gamma) = H(X_\alpha, X_\gamma) - H(X_\gamma). \quad (4)$$

For instance

$$H(X_1 | X_2) = [0, -1, 1, 0, 0, 0] \cdot h.$$
ii. Show that

\[ I(X_\alpha; X_\beta | X_\gamma) \]

can also be written as a linear combination of \( h \).

**Solution:**

Use the identity

\[ I(X_\alpha; X_\beta | X_\gamma) = H(X_\alpha | X_\gamma) + H(X_\beta | X_\gamma) - H(X_\alpha, X_\beta | X_\gamma) \]

and the identity from 2(a)i to write a mutual information measure (conditional and unconditional) as linear combination of unconditional entropies.

The representation by \( h \) is *unique* and called *canonical form*.

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

**Solution:**

Using the result from 2(a)i and 2(a)ii, every information measure can be expressed by linear combination of *unconditional* joint entropies. Hence, a linear combination of information measures can also be represented by \( h^\top h \).

For instance, the corresponding \( f \) for \( I(X_1; X_2, X_3) - I(X_1; X_2 | X_3) \) is:

\[
I(X_1; X_2, X_3) = H(X_1) + H(X_2, X_3) - H(X_1, X_2, X_3)
\]
\[
\rightarrow f_1 = [1, 0, 0, 0, 0, 1, -1]^\top
\]
\[
I(X_1; X_2 | X_3) = H(X_1 | X_3) + H(X_2 | X_3) - H(X_1, X_2 | X_3)
\]
\[
\rightarrow f_2 = [0, 0, 0, -1, 1, 0, 0]^\top
\]
\[
H(X_1 | X_3) = H(X_1, X_3) - H(X_3)
\]
\[
\rightarrow f_3 = [0, 0, 0, -1, 0, 1, 0]^\top
\]
\[
H(X_2 | X_3) = H(X_2, X_3) - H(X_3)
\]
\[
\rightarrow f_4 = [0, 0, 0, -1, 0, 0, 1]^\top
\]
\[
I(X_1; X_2, X_3) - I(X_1; X_2 | X_3) = (f_1 - f_2 - f_3 + f_4)^\top h
\]
\[
\rightarrow f = [1, 0, 0, 1, -1, 0, 0]^\top
\]

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

\[
H(X_1 | X_2, X_3) \geq 0, \ I(X_1; X_2) \geq 0, \ I(X_1; X_2 | X_3) \geq 0,
\]
\[
H(X_2 | X_1, X_3) \geq 0, \ I(X_2; X_3) \geq 0, \ I(X_2; X_3 | X_1) \geq 0,
\]
\[
H(X_3 | X_2, X_3) \geq 0, \ I(X_1; X_3) \geq 0, \ I(X_1; X_3 | X_2) \geq 0.
\]
Find a matrix, $G : 9 \times 7$, such that $G h \geq 0$ if and only if all elemental inequalities hold.

**Solution:**

We write each elemental inequality in as a row vector $f^\top \text{s.t } f^\top h \geq 0$.

Matrix $G$ consist of all elemental inequalities. For $n = 3$,

$$G = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & -1 & 0 & 0 & 0 & 1 \\
1 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 1 & -1 \\
1 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 1 & -1 \\
0 & 1 & 0 & 1 & 0 & -1 & 0 \\
-1 & 0 & 1 & 0 & 1 & 0 & -1 \\
\end{bmatrix} \quad (21)$$

Each element in $G h \geq 0$ corresponds to an inequality.

(d) A linear programming optimization problem has the form:

$$x^* = \min_x f^\top x \quad (22a)$$

$$\text{Subject to } A x \leq b \quad (22b)$$

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

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

**Solution:**

If an information inequality $f^\top h \geq 0$ always holds, the minimum of the expression is zero.

We can check what is the minimum of the expression, where vector $h$ is the variables vector $x$,

which must satisfy $-G x \leq 0$. Therefore, we identify $A$ as $-G$, $x$ as $h$ and $f$ is the vector that represents the inequality in the minimization problem.

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

**Solution:** Assume that the minimum of $f^\top h$ is zero. Then, due to non-negativity constraints on entropies, the expression is greater than or equal to zero. Since
all entropies that are induced from probability mass functions, must hold non-

negativity constraints, then the expression must also be greater than or equal to

zero. Mathematically,

\[
\min_{P_{X_1,X_2,X_3}} f^\top h(P_{X_1,X_2,X_3}) \geq \min_{-\mathbf{G}h \leq 0} f^\top h \\
\geq 0
\]

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

Solution: If the minimum value of the expression is less than zero, we cannot

conclude about the inequality. There may be other constraints on information

measures, that we didn’t consider, such that minimum value will change according

to them.

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

- \(I(X_1; X_2, X_3) - I(X_1; X_2|X_3) \geq 0\),
- \(H(X_1) + H(X_2) - H(X_1, X_2) \geq 0\),
- \(H(X_1, X_2) - H(X_1|X_2) \geq 0\).

Solution: The minimum value of the expression \(I(X_1; X_2, X_3) - I(X_1; X_2|X_3)\) is

zero, since \(I(X_1; X_2, X_3) - I(X_1; X_2|X_3) = I(X_1; X_3) \geq 0\).

We write an algorithm which minimizes the expression, subject to the elemental

information inequalities; if the returned value is zero, then the expression always

holds.

Algorithm:
\[
\mathbf{G} = \begin{bmatrix} 0, 0, 0, 0, 0, -1, 1; \\
0, 0, 0, 0, -1, 0, 1; \\
0, 0, -1, 0, 0, 0, 1; \\
1, 1, -1, 0, 0, 0, 0; \\
0, 0, 0, -1, 1, 1, -1; \\
1, 0, 0, 1, -1, 0, 0; \\
0, -1, 1, 0, 0, 1, -1; \\
0, 1, 0, 1, 0, -1, 0; \\
-1, 0, 1, 0, 1, 0, -1 \end{bmatrix}; \\
b = \text{zeros}(9,1); \\
f = [1, 0, 0, 1, -1, 0, 0]'; \\
h = \text{linprog}(f, -\mathbf{G}, b); \\
\text{if abs(f'*h) <= 1e-8; display('Always holds'); end;}
\]

Note that the default tolerance for \texttt{linprog} returned value is \(10^{-8}\).
Similarly for other inequalities.

(e) 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
\[
q^\top h = I(X_1; X_3|X_2) \quad (23)
\]

Solution: We use the canonical form of \(I(X_1; X_3|X_2)\) in order to find \(q\).

\[
I(X_1; X_3|X_2) = I(X_1, X_2; X_3) - I(X_2; X_3)
= H(X_1, X_2) + H(X_3) - H(X_1, X_2, X_3) - (H(X_2) + H(X_3) - H(X_2, X_3))
= -H(X_2) + H(X_1, X_2) - H(X_3) + H(X_2, X_3)
\]

Therefore, \(q^\top = [0, -1, 1, -1, 0, 1, 0]\).

ii. Linear optimization methods (as in 22) can also account for equality constraints:
\[
x^* = \min_x f^\top x \quad (24a)
\]
\[
\text{Subject to: } Ax \leq b \quad (24b)
\]
\[
Cx = d \quad (24c)
\]

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 \(X_1 - X_2 - X_3\)?

Solution: Similarly to matrix \(A\), \(C = q^\top\). The expression equals to zero, therefore \(d = [0]\).

iii. Write a MATLAB code (use the command \texttt{linprog}(f,A,b,C,d)) to proves that \(I(X_1; X_2) - I(X_1; X_3) \geq 0\).

Solution: First, we need to find the canonical form of \(I(X_1; X_2) - I(X_1; X_3)\):
\[
I(X_1; X_2) - I(X_1; X_3) = H(X_2) - H(X_1, X_2) - H(X_3) + H(X_2, X_3)
\]

Therefore, \(f^\top = [0, 1, -1, -1, 0, 1, 0]\).

Algorithm:
\[
G=[0,0,0,0,0,-1,1; \ldots \\
0,0,0,0,-1,0,1; \ldots \\
0,0,-1,0,0,0,1; \ldots \\
1,1,-1,0,0,0,0; \ldots \\
0,0,0,-1,1,1,-1; \ldots 
\]
\[1,0,0,1,-1,0,0;\ldots\]
\[0,-1,1,0,0,1,-1;\ldots\]
\[0,1,0,1,0,-1,0;\ldots\]
\[-1,0,1,0,1,0,-1]\]
\[b=\text{zeros}(9,1);\]
\[C= [0, -1, 1, -1, 0, 1, 0];\]
\[d=0;\]
\[f=[0, 1, -1, -1, 0, 1, 0]';\]
\[h=\text{linprog}(f,-G,b,C,d);\]
\[\text{if abs}(f'*h)\leq 1e-8; \text{display('Inequality holds'); end;}\]