ECE 563 FA25 HW1 Solutions

Problem 1. (An axiomatic characterization of the Rényi entropy.)

Solution (A sketch of solution)**.** See the original paper by Rényi [1]. The main idea is to use some results about quasi-arithmetic means (such as [2, Theorem 83]) to force the generator ϕ to be the ones for the Rényi entropy.

Remark. The terminology in [1] is somewhat different from the description of this problem. In particular, [1] considered the measures of randomness to be defined for "generalized probability distributions", which means non-negative numbers whose sum is less than or equal to one. You may try figuring out how the axioms in [1] are related to the ones in this problem.

Problem 2. Show that

$$H(X,Y) + H(Y,Z) - H(Y) = H(X,Y,Z) + I(X;Z|Y). \tag{2.1}$$

Solution. By the definition of conditional mutual information, we have

$$I(X;Z|Y) = H(X|Y) - H(X|Y,Z). \tag{2.2}$$

On the other hand, by the chain rule of conditional entropy, we have

$$H(X,Y,Z) = H(Y,Z) + H(X|Y,Z). \tag{2.3}$$

Summing up (2.2) and (2.3) yields

$$H(X,Y,Z) + I(X;Z|Y) = H(X|Y) + H(Y,Z). \tag{2.4}$$

At the same time, we have by the chain rule of conditional entropy again that

$$H(X|Y) = H(X,Y) - H(Y). \tag{2.5}$$

Putting (2.5) into (2.4), we get

$$H(X,Y,Z) + I(X;Z|Y) = H(X,Y) - H(Y) + H(Y,Z),$$

which is exactly (2.1).

Problem 3. (Han's Theorem.) Show that

$$\frac{1}{n-1}\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}|X_i) \le H(X_1,\dots,X_n) \le \frac{1}{n-1}\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}). \tag{3.1}$$

Remark. Here $[n]$ denotes the set $\{1,2,\dots,n\}$.

Solution. We first prove the first inequality

$$\frac{1}{n-1}\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}|X_i) \le H(X_1,\dots,X_n). \tag{3.2}$$

To start with, note that by the chain rule of conditional entropy, we have for each $i \in [n]$ that

$$H(X_{[n]\setminus\{i\}}|X_i) = H(X_1,\dots,X_n) - H(X_i). \tag{3.3}$$

Summing up (3.3) over all $i \in [n]$, we get

$$\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}|X_i) = nH(X_1,\dots,X_n) - \sum_{i=1}^{n} H(X_i). \tag{3.4}$$

Then, note that by the chain rule and the fact that conditioning reduces entropy, we have

$$\begin{aligned} H(X_1,\dots,X_n) &= \sum_{i=1}^{n} H(X_i|X_{i-1},\dots,X_1) \\ &\le \sum_{i=1}^{n} H(X_i). \end{aligned} \tag{3.5}$$

Putting (3.5) into (3.4) yields

$$\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}|X_i) \leq nH(X_1,\dots,X_n) - H(X_1,\dots,X_n)$$
$$= (n-1)H(X_1,\dots,X_n). \tag{3.6}$$

Dividing both sides of (3.6) by $n-1$ gives (3.2).

We now prove the second inequality

$$H(X_1,\dots,X_n) \leq \frac{1}{n-1}\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}) \tag{3.7}$$

By the chain rule, we have for each $i \in [n]$ that

$$H(X_{[n]\setminus\{i\}}) = H(X_1,\dots,X_n) - H(X_i|X_{[n]\setminus\{i\}}). \tag{3.8}$$

Summing up (3.8) over all $i \in [n]$, we obtain

$$\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}) = nH(X_1,\dots,X_n) - \sum_{i=1}^{n} H(X_i|X_{[n]\setminus\{i\}}). \tag{3.9}$$

Now, notice that by, again, the chain rule and the fact that conditioning reduces entropy, we have

$$\sum_{i=1}^{n} H(X_i|X_{[n]\setminus\{i\}}) \leq \sum_{i=1}^{n} H(X_i|X_{i-1},\dots,X_1)$$
$$= H(X_1,\dots,X_n). \tag{3.10}$$

Putting (3.10) into (3.9) gives

$$\sum_{i=1}^{n} H(X_{[n]\setminus\{i\}}) \geq nH(X_1,\dots,X_n) - H(X_1,\dots,X_n)$$
$$= (n-1)H(X_1,\dots,X_n). \tag{3.11}$$

Dividing both sides of (3.11) by $n-1$ yields (3.7).

Finally, combining (3.2) and (3.7) gives (3.1).

Problem 4. Prove that

$$I(X_1,...,X_n;Y) = \sum_{i=1}^{n} I(X_i;Y|X_1,...,X_{i-1}). \tag{4.1}$$

Then, prove the tensorization inequality for mutual information: Let $(X_1,Y_1),\dots,(X_n,Y_n)$ be random pairs satisfying the following property

$$p_{Y|X}(y_1,\dots,y_n|x_1,\dots,x_n) = \prod_{i=1}^{n} p_{Y_i|X_i}(y_i|x_i), \tag{4.2}$$

where we define $X = (X_1,\dots,X_n)$ and $Y = (Y_1,\dots,Y_n)$. Show that

$$I(X;Y) \leq \sum_{i=1}^{n} I(X_i;Y_i). \tag{4.3}$$

Remark. Without the additional assumption in (4.2), the inequality in (4.3) does not hold for general $2n$ random variables $X_1,\dots,X_n$ and $Y_1,\dots,Y_n$. You are encouraged to find four random variables X_1,X_2,Y_1,Y_2 such that $I(X_1,X_2;Y_1,Y_2) > I(X_1;Y_1) + I(X_2;Y_2)$.

Solution. The proof of (4.1) can already be seen on Cover and Thomas [3, Page 24].

As for (4.3), we start with the decomposition $I(X;Y) = H(Y) - H(Y|X)$. Then, note that

$$H(Y) \leq \sum_{i=1}^{n} H(Y_i), \tag{4.4}$$

the derivation of which is exactly the same as (3.5). At the same time, we have

$$\begin{aligned} H(Y|X) &= -\sum_{x_1,\dots,x_n,y_1,\dots,y_n} p_{X,Y}(x_1,\dots,x_n,y_1,\dots,y_n)\log p_{Y|X}(y_1,\dots,y_n|x_1,\dots,x_n) \\ &= -\sum_{x_1,\dots,x_n,y_1,\dots,y_n} p_{X,Y}(x_1,\dots,x_n,y_1,\dots,y_n)\log \prod_{i=1}^{n} p_{Y|X}(y_i|x_i) \qquad (4.5) \\ &= -\sum_{x_1,\dots,x_n,y_1,\dots,y_n} p_{X,Y}(x_1,\dots,x_n,y_1,\dots,y_n)\sum_{i=1}^{n}\log p_{Y_i|X_i}(y_i|x_i) \\ &= -\sum_{i=1}^{n}\sum_{x_1,\dots,x_n,y_1,\dots,y_n} p_{X,Y}(x_1,\dots,x_n,y_1,\dots,y_n) p_{Y_i|X_i}(y_i|x_i) \\ &= -\sum_{i=1}^{n}\sum_{x_i,y_i} p_{X_i,Y_i}(x_i,y_i)\log p_{Y_i|X_i}(y_i|x_i) \\ &= \sum_{i=1}^{n} H(Y_i|X_i), \qquad (4.6) \end{aligned}$$

where in (4.5) we used the additional assumption in (4.2). Combining (4.4) and (4.6), we get

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &= \sum_{i=1}^{n}(H(Y_i) - H(Y_i - X_i)) \\ &= \sum_{i=1}^{n} I(X_i;Y_i), \end{aligned}$$

which proves (4.3).

Problem 5. [3, Problem 2.9], Metric.

Solution.

(a) We check the following four properties:

- Since conditional entropy is non-negative, we have $\rho(X,Y) \ge 0$.
- Since addition is commutative in real numbers, we have $\rho(X,Y) = \rho(Y,X)$.
- By the non-negativity of conditional entropy again, $\rho(X,Y) = 0$ implies $H(X|Y) = H(Y|X) = 0$. Note that $H(X|Y) = 0$ means that X is a function of Y, and similarly $H(Y|X) = 0$ implies that Y is a function of X. Therefore, from $H(X|Y) = H(Y|X) = 0$ we can infer that there is a bijection between X and Y, or equivalently $X = Y$ according to the notation defined in the problem description. Conversely, it is clear that if there is a one-to-one correspondence between X and Y, then $\rho(X,Y) = H(X|Y) + H(Y|X) = 0 + 0 = 0$. As a result, $\rho(X,Y) = 0$ if and only if $X = Y$.
- Note that

$$H(X|Y) = H(X|Y,Z) + I(X;Z|Y), \qquad (5.1)$$
$$H(Y|Z) = H(Y|X,Z) + I(X;Y|Z), \qquad (5.2)$$
$$H(X|Z) = H(X|Y,Z) + I(X;Y|Z). \qquad (5.3)$$

Therefore, from (5.1), (5.2), and (5.3), we obtain

$$\begin{aligned} H(X|Y) + H(Y|Z) - H(X|Z) &= H(Y|X,Z) + I(X;Z|Y) \qquad (5.4) \\ &\ge 0, \qquad (5.5) \end{aligned}$$

since both conditional entropy and conditional mutual information are non-negative. A similar derivation gives

$$H(Y|X) + H(Z|Y) - H(Z|X) \ge 0. \qquad (5.6)$$

Summing up (5.5) and (5.6) gives $\rho(X,Y) + \rho(Y,Z) \ge \rho(Z,X)$.

These arguments show that ρ is a metric.

Remark. The equality in (5.4) can be easily visualized on a Venn diagram. A takeaway is: Whenever you want to simplify an expression involving (conditional) entropy and/or mutual information of three random variables, you may first use a Venn diagram to guess the final answer and then prove it.

(b) We first expand

$$\begin{aligned}\rho(X,Y) &= H(X,Y) - H(Y) + H(X,Y) - H(X) \\ &= 2H(X,Y) - H(X) - H(Y), \end{aligned} \tag{5.7}$$

which already proves the third expression. Then, from (5.7) and the relationship $I(X;Y) = H(X) + H(Y) - H(X,Y)$, we have

$$\begin{aligned}\rho(X,Y) &= H(X,Y) + H(X,Y) - H(X) - H(Y) \\ &= H(X,Y) - I(X;Y), \end{aligned} \tag{5.8}$$

which is the second expression. Lastly, from (5.8) and the relationship $H(X;Y) = H(X) + H(Y) - I(X;Y)$, we have

$$\begin{aligned}\rho(X,Y) &= H(X) + H(Y) - I(X;Y) - I(X;Y) \\ &= H(X) + H(Y) - 2I(X;Y), \end{aligned}$$

which is the first expression.

Problem 6. [3, Problem 2.10] Entropy of a disjoint mixture.

Solution.

(a) Note that the PMF of X, denoted as $p(\cdot)$, satisfy

$$p(x) = \begin{cases} \alpha p_1(x), & \text{if } x \in \mathcal{X}_1, \\ (1-\alpha)p_2(x), & \text{if } x \in \mathcal{X}_2. \end{cases}$$

It follows that the entropy of X is

$$\begin{aligned} H(X) &= -\sum_{x=1}^{m} p(x)\log p(x) - \sum_{x=m+1}^{n} p(x)\log p(x) \\ &= -\sum_{x=1}^{m} \alpha p_1(x) \log \alpha p_1(x) - \sum_{x=m+1}^{n} (1-\alpha)p_2(x)\log(1-\alpha)p_2(x) \\ &= -\alpha \sum_{x=1}^{m} p_1(x)(\log\alpha + \log p_1(x)) - (1-\alpha) \sum_{x=m+1}^{n} p_2(x)(\log(1-\alpha) + \log p_2(x)) \\ &= -\alpha\log\alpha + \alpha H(X_1) - (1-\alpha)\log(1-\alpha) + (1-\alpha)H(X_2) \\ &= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2), \end{aligned}$$

where $H(x) \coloneqq -x\log x - (1-x)\log(1-x)$ denotes the binary entropy.
An alternative solution to this subproblem is as follows: Define the random variable

$$B = \begin{cases} 1, & \text{if } X \in \mathcal{X}_1, \\ 2, & \text{if } X \in \mathcal{X}_2. \end{cases}$$

In particular, B is a function of X. Furthermore, B follows a distribution of a shifted Bernoulli(α) random variable. Also note that given $B = 1$, the distribution of X is X_1, and similarly, given $B = 2$, X follows the distribution of X_2. Therefore, we have

$$\begin{aligned} H(X) &= H(X,B) \\ &= H(B) + H(X|B) \\ &= H(\alpha) + \mathbb{P}(B=1)H(X|B=1) + \mathbb{P}(B=2)H(X|B=2) \\ &= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2), \end{aligned}$$

which gives the same result as the first solution.

(b) Let $f(\alpha) \coloneqq H(X) = H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2)$. By a direct calculation, it can be shown that

$$f'(\alpha) = \log(1-\alpha) - \log(\alpha) + H(X_1) - H(X_2), \tag{6.1}$$

$$f''(\alpha) = -\frac{1}{\ln 2}\left(\frac{1}{1-\alpha} + \frac{1}{\alpha}\right) < 0, \tag{6.2}$$

where $\ln(\cdot)$ denotes the natural log. Solving $f'(\alpha) = 0$ in (6.1) gives

$$\alpha = \frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}}. \tag{6.3}$$

Furthermore, (6.2) implies that f attains maximum at $\alpha = \frac{2^{H(X_1)}}{2^{H(X_1)}+2^{H(X_2)}}$, as derived in (6.3). Then, a direct calculation shows that

$$f(\frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}}) = \log(2^{H(X_1)} + 2^{H(X_2)}). \tag{6.4}$$

It follows that for any $\alpha \in [0,1]$ we have

$$H(\alpha) \leq \log(2^{H(X_1)} + 2^{H(X_2)}). \tag{6.5}$$

Or equivalently, $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$.

Interpretation: $2^{H(X_1)}$ is the "effective alphabet size" of X_1, and $2^{H(X_2)}$ is the "effective alphabet size" of X_2. At the same time, these two "effective alphabets" are still disjoint. Therefore, X is a random variable over an "effective alphabet" of size $2^{H(X_1)} + 2^{H(X_2)}$. It follows that the entropy of X cannot exceed the log of the "effective alphabet size", which is exactly saying that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$.

Problem 7. [3, Problem 2.11] A measure of correlation.

Solution.

(a) Using the relation $I(X_1; X_2) = H(X_1) - H(X_1|X_2)$, we have

$$\rho = \frac{H(X_1) - H(X_1|X_2)}{H(X_1)} = \frac{I(X_1;X_2)}{H(X_1)}.$$

(b) From the definition of ρ and the fact that (conditional) entropy is non-negative, we have $\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} \leq 1 - 0 = 1$. On the other hand, from Part (a) and the fact that mutual information is non-negative, we have $\rho \geq 0$.

(c) From Part (a) and (b), we can see that $\rho = 0$ if and only if $I(X_1; X_2) = 0$. That is, $\rho = 0$ if and only if X_1 and X_2 are independent.

(d) From Part (a) and (b), we can also see that $\rho = 1$ if and only if $H(X_2|X_1) = 0$. That is, $\rho = 1$ if and only if X_2 is a function of X_1. At the same time, since X_1 and X_2 have the same distribution, we have from Part (a) that

$$\begin{aligned}\rho &= \frac{I(X_1;X_2)}{H(X_2)} \\ &= 1 - \frac{H(X_1|X_2)}{H(X_2)}.\end{aligned}$$

It follows that $\rho = 1$ if and only if X_1 is a function of X_2.

We can conclude that the following statements are equivalent (that is, one implies the other two):

- $\rho = 0$,
- X_1 is a function of X_2,
- X_2 is a function of X_1.

Problem 8. [3, Problem 2.25] Venn diagrams.

Solution. We first find X, Y, Z such that $I(X;Y;Z) = I(X;Y) - I(X;Y|Z) < 0$.

One way to find such counterexamples is to construct pairwise independent but not mutually independent random variables. Consider X and Y to be i.i.d. Bernoulli($\frac{1}{2}$) random variables. That is, $\mathbb{P}(X=0) = \mathbb{P}(X=1) = \mathbb{P}(Y=0) = \mathbb{P}(Y=1) = \frac{1}{2}$, and X is independent of Y. In particular, since they are independent, we have

$$I(X;Y) = 0. \tag{8.1}$$

Then, define

$$Z = \begin{cases} 1, & \text{if } X = Y, \\ 0, & \text{if } X \neq Y. \end{cases}$$

It can be checked that:

1) X and Z are independent.
2) X is a function of Y and Z.

It follows that

$$H(X|Z) = H(X) = 1, \tag{8.2}$$

and that

$$H(X|Y,Z) = 0. \tag{8.3}$$

Combining (8.2) and (8.3) gives

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= 1. \end{aligned} \tag{8.4}$$

From (8.1) and (8.4), we have

$$\begin{aligned} I(X;Y;Z) &= I(X;Y) - I(X;Y|Z) \\ &= 0 - 1 \\ &= -1 \\ &< 0. \end{aligned}$$

(a) Using the following equalities

$$\begin{aligned} H(X|Z) &= H(X) - I(X;Z), \\ H(Y|Z) &= H(Y) - I(Y;Z), \\ H(X,Y|Z) &= H(X,Y,Z) - H(Z), \end{aligned}$$

we can obtain

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) + H(Y|Z) - H(X,Y|Z) \\ &= -H(X,Y,Z) + H(X) + H(Y) + H(Z) - I(X;Z) - I(Y;Z). \end{aligned} \tag{8.5}$$

Then, from (8.5) we have

$$\begin{aligned} I(X;Y;Z) &= I(X;Y) - I(X;Y|Z) \\ &= H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(X;Z). \end{aligned}$$

(b) Using $I(X;Y) = H(X) + H(Y) - H(X,Y)$, $I(Y;Z) = H(Y) + H(Z) - H(Y,Z)$, and $I(X;Z) = H(X) + H(Z) - H(X,Z)$, from Part (a) we have

$$\begin{aligned} I(X;Y;Z) &= H(X,Y,Z) - H(X) - H(Y) - H(Z) + 2H(X) + 2H(Y) + 2H(Z) - H(X,Y) - H(Y,Z) - H(X,Z) \\ &= H(X,Y,Z) - H(X,Y) - H(Y,Z) - H(X,Z) + H(X) + H(Y) + H(Z). \end{aligned}$$

References

[1] A. Rényi, “On measures of entropy and information,” in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, vol. 4. University of California Press, 1961, pp. 547–562.
[2] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge university press, 1952.
[3] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.