EEE 551 Information Theory (Spring 2022)

Homework 1 Solutions

---

1. *Example of joint entropy.* Let $p(x,y)$ be given by

		X		
		0	1	2
Y	0	$\frac{1}{3}$	$\frac{1}{6}$	0
	1	0	$\frac{1}{6}$	$\frac{1}{3}$

Find:

(a) $H(X)$, $H(Y)$.
(b) $H(X|Y)$, $H(Y|X)$.
(c) $H(X,Y)$.
(d) $H(Y) - H(Y|X)$.
(e) $I(X;Y)$.
(f) Draw a Venn diagram for the quantities in parts (a) through (e).

Solution:

(a) The marginal distribution of X is $p_X = [1/3, 1/3, 1/3]$. Thus, X is uniform over the set $\{0, 1, 2\}$, so $H(X) = \log 3$. The marginal distribution of Y is $[1/2, 1/2]$. Thus, Y is uniform over the set $\{0, 1\}$, so $H(Y) = \log 2 = 1$.

(b) $H(X|Y) = \frac{1}{2}H(X|Y=0) + \frac{1}{2}H(X|Y=1) = \frac{1}{2}H(2/3, 1/3) + \frac{1}{2}H(1/3, 2/3) = H(1/3, 2/3) = \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2} = \log 3 - \frac{2}{3} = 0.9183$.

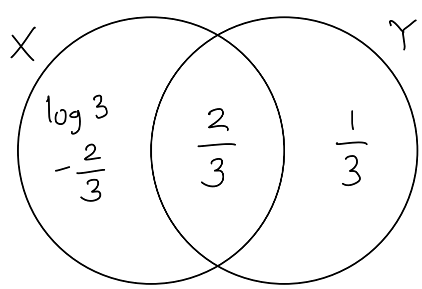
$H(Y|X) = \frac{1}{3}H(Y|X=0) + \frac{1}{3}H(Y|X=1) + \frac{1}{3}H(Y|X=2) = \frac{1}{3}(0) + \frac{1}{3}(1) + \frac{1}{3}H(0) = \frac{1}{3}$.

(c) By the chain rule, $H(X,Y) = H(X) + H(Y|X) = \log 3 + \frac{1}{3}$. As a check, we can use the chain rule the other way: $H(X,Y) = H(Y) + H(X|Y) = 1 + (\log 3 - \frac{2}{3}) = \log 3 + \frac{1}{3}$.

(d) $H(Y) - H(Y|X) = 1 - \frac{1}{3} = \frac{2}{3}$.

(e) $I(X;Y) = H(Y) - H(Y|X) = \frac{2}{3}$.

(f) The following Venn diagram shows the information measures for this example:



2. Problem 2.4 from Cover-Thomas: *Entropy of functions of a random variable.* Let X be a discrete random variable. Show that the entropy of a function of X is less than or equal to the entropy of X by justifying the following steps:

$$\begin{aligned} H(X, g(X)) &\stackrel{\textbf{(a)}}{=} H(X) + H(g(X)|X) \\ &\stackrel{\textbf{(b)}}{=} H(X), \\ H(X, g(X)) &\stackrel{\textbf{(c)}}{=} H(g(X)) + H(X|g(X)) \\ &\stackrel{\textbf{(d)}}{\geq} H(g(X)). \end{aligned}$$

Thus, $H(g(X)) \leq H(X)$.

Solution:

(a) Follows from the chain rule.

(b) Follows because $H(Y|X) = 0$ when Y is a function of X.

(c) Follows from the chain rule.

(d) Follows because conditional entropy is non-negative.

3. *Yes/No question.* Let X be a random variable with PMF given by

x	a	b	c	d	e	f	g
$p(x)$	1/4	1/4	1/8	1/8	1/8	1/16	1/16

Let Y be a random variable with PMF given by

y	a	b	c	d
$p(y)$	0.4	0.3	0.2	0.1

(a) Devise a strategy to determine X by a series of Yes/No questions such that the expected number of questions is exactly equal to $H(X)$. Each question may depend on the outcome of the previous question.

(b) For Y, find a strategy of Yes/No questions to minimize the expected number of questions. What is the smallest possible expected number of questions? In this case, how does it compare to $H(Y)$?

Solution:

(a) Consider the following Yes/No question strategy:

Is $X \in \{a, b\}$?

- If Yes: Is $X = a$?
  - If Yes, then $X = a$
  - If No, then $X = b$
- If No: Is $X \in \{c, d\}$?
  - If Yes: Is $X = c$?
    - If Yes, then $X = c$

    * If No, then $X = d$
  - If No: Is $X = e$?
    * If Yes, then $X = e$
    * If No: Is $X = f$?
      - If Yes, then $X = f$
      - If No, then $X = g$

For this strategy, the number of questions asked for each X value is given in the following table:

x	a	b	c	d	e	f	g
Number of questions	2	2	3	3	3	4	4

Thus the expected number of questions is

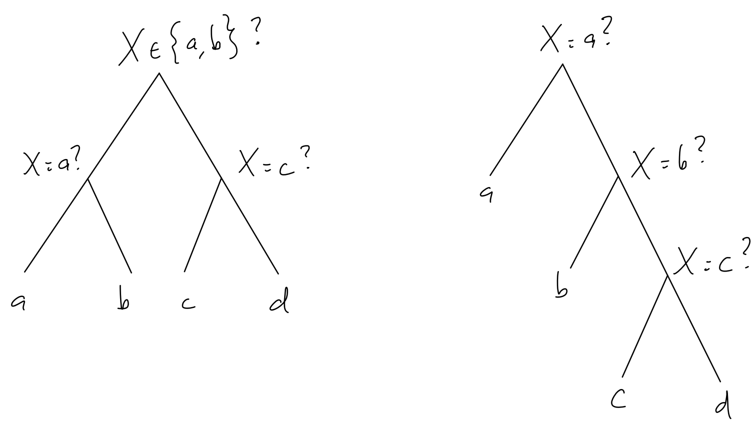
$$\frac{1}{4}\cdot 2+\frac{1}{4}\cdot 2+\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3+\frac{1}{16}\cdot 4+\frac{1}{16}\cdot 4=2.625.$$

Meanwhile, the entropy is

$$H(X)=\frac{1}{4}\log 4+\frac{1}{4}\log 4+\frac{1}{8}\log 8+\frac{1}{8}\log 8+\frac{1}{8}\log 8+\frac{1}{16}\log 16+\frac{1}{16}\log 16=2.625.$$

Thus, the expected number of questions is equal to the entropy.

(b) The two reasonable strategies for guessing Y are illustrated below:



The strategy on the left has an expected length of 2 (since all possibilities involve 2 questions). The strategy on the right has an expect length of:

$$0.4(1)+0.3(2)+0.2(3)+0.1(3)=1.9.$$

Thus, the strategy on the right minimizes the expected number of questions. The entropy is

$$H(Y)=-0.4\log 0.4-0.3\log 0.3-0.2\log 0.2-0.1\log 0.1=1.8464.$$

Thus, the expected number of questions is larger than the entropy.

4. *Mutual information vs. conditional mutual information.* In general, conditioning neither decreases nor increases mutual information; i.e. either the mutual information $I(X;Y)$ or the conditional mutual information $I(X;Y|Z)$ can be larger. This problem shows that under certain conditions, one can conclude that one of these quantities is larger than the other. Prove the following two statements:

   (a) If $X \to Y \to Z$, then $I(X;Y|Z) \leq I(X;Y)$.

   (b) If X and Z are independent, then $I(X;Y) \leq I(X;Y|Z)$.

   Hint: Expand $I(X;Y,Z)$ using the chain rule in two different ways.

   Solution: For both parts, we will use the following equality, which is derived from the chain rule for mutual information:

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

   (a) If $X \to Y \to Z$, then $I(X;Z|Y) = 0$. Thus, using the equality from above,

$$I(X;Y|Z) = I(X;Y) + I(X;Z|Y) - I(X;Z) = I(X;Y) - I(X;Z) \leq I(X;Y).$$

   (b) If X and Z are independent, then $I(X;Z) = 0$. Thus, again using the equality from above,

$$I(X;Y) = I(X;Z) + I(X;Y|Z) - I(X;Z|Y) = I(X;Y|Z) - I(X;Z|Y) \leq I(X;Y|Z).$$

5. Problem 2.27 from Cover-Thomas: *Grouping rule for entropy.* Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ be a probability distribution on m elements (i.e., $p_i \geq 0$ and $\sum_{i=1}^m p_i = 1$). Define a new distribution $\mathbf{q}$ on $m-1$ elements as $q_1 = p_1, q_2 = p_2, \ldots, q_{m-2} = p_{m-2}$, and $q_{m-1} = p_{m-1} + p_m$ [i.e., the distribution $\mathbf{q}$ is the same as $\mathbf{p}$ on $\{1, 2, \ldots, m-2\}$, and the probability of the last element in $\mathbf{q}$ is the sum of the last two probabilities of $\mathbf{p}$]. Show that

$$H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\right).$$

   Solution: Let $X \in \{1, \ldots, m\}$ be a random variable with distribution $\mathbf{p}$, i.e. $p_X(i) = p_i$. Thus $H(X) = H(\mathbf{p})$. Define a random variable $Y \in \{1, \ldots, m-1\}$ as

$$Y = \begin{cases} X, & X \leq m-1 \\ m-1, & X = m. \end{cases}$$

   Note that Y is distributed according to $\mathbf{q}$. Moreover, Y is a function of X, so $H(Y|X) = 0$. Thus

$$H(\mathbf{p}) = H(X) = H(X) + H(Y|X) = H(X,Y) = H(Y) + H(X|Y) = H(\mathbf{q}) + \sum_{i=1}^{m-1} q_i H(X|Y=i).$$

   For all $i < m-1$, if $Y = i$ then $X = i$. Thus for $i < m-1$, $H(X|Y=i) = 0$. Therefore the last term above is simply $q_{m-1} H(X|Y=m-1)$. Recall $q_{m-1} = p_{m-1} + p_m$. Given $Y = m-1$, $X = m-1$ with probability $\frac{p_{m-1}}{p_{m-1}+p_m}$ and $X = m$ with probability $\frac{p_m}{p_{m-1}+p_m}$. Thus $H(X|Y=m-1) = H\left(\frac{p_{m-1}}{p_{m-1}+p_m}, \frac{p_m}{p_{m-1}+p_m}\right)$. Putting this together we find

$$H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\right).$$

6. *Rényi Entropy.* Rényi entropy is a different way of defining an "entropy" that generalizes Shannon's measure. Rényi entropy has a parameter α, which can be any positive number except 1. Rényi entropy is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left[\sum_{x \in \mathcal{X}} p(x)^\alpha \right].$$

(a) Find $H_\alpha(X)$ if X is a uniform random variable with an alphabet of size m. Compare your answer to the corresponding value for the standard entropy.

(b) Plot $H_\alpha(X)$ as a function of α where X is Bern(0.1).

(c) Prove that $\lim_{\alpha \to 1} H_\alpha(X) = H(X)$, where $H(X)$ is the standard entropy. For this reason, the standard entropy is sometimes written $H_1(X)$. *Hint:* Use L'Hôpital's rule.

(d) If X and Y are independent, prove that $H_\alpha(X, Y) = H_\alpha(X) + H_\alpha(Y)$.

(e) Show that Rényi entropy does *not* satisfy the grouping rule, as defined in Problem 5. To do this, find a distribution $\mathbf{p}$ where the formula in Problem 5 is violated for Rényi entropy.

Solution:

(a) The Rényi entropy for a uniform variable is given by

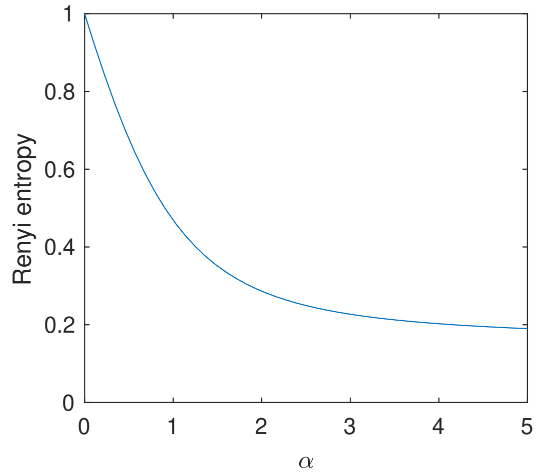
$$\begin{aligned}
H_\alpha(X) &= \frac{1}{1-\alpha} \log \left[\sum_{x \in \mathcal{X}} \left(\frac{1}{m} \right)^\alpha \right] \\
&= \frac{1}{1-\alpha} \log \left[\frac{m}{m^\alpha} \right] \\
&= \frac{1}{1-\alpha} \log \left[m^{1-\alpha} \right] \\
&= \frac{1-\alpha}{1-\alpha} \log m \\
&= \log m.
\end{aligned}$$

This is identical to the standard entropy value.

(b) The Rényi entropy of a Bern(0.1) random variable is

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left[(0.1)^\alpha + (0.9)^\alpha \right].$$

As a function of α, this is plotted as follows:



(c) The limit is computed as follows:

$$\begin{aligned}
\lim_{\alpha\to 1} H_\alpha(X) &= \lim_{\alpha\to 1} \frac{\ln\left[\sum_{x\in\mathcal{X}} p(x)^\alpha\right]}{(1-\alpha)\ln 2} \\
&\overset{(a)}{=} \lim_{\alpha\to 1} \frac{\frac{d}{d\alpha}\ln\left[\sum_{x\in\mathcal{X}} p(x)^\alpha\right]}{\frac{d}{d\alpha}(1-\alpha)\ln 2} \\
&= \lim_{\alpha\to 1} \frac{\frac{\sum_x p(x)^\alpha \ln p(x)}{\sum_x p(x)^\alpha}}{-\ln 2} \\
&= \frac{-\sum_x p(x)\log p(x)}{\sum_x p(x)} \\
&\overset{(b)}{=} -\sum_x p(x)\log p(x) \\
&= H(X)
\end{aligned}$$

where in (a) we have used L'Hôpital's rule, and in (b) we have used the fact that $\sum_{x\in\mathcal{X}} p(x) = 1$.

(d) Assume X and Y are independent, so $p(x,y) = p(x)p(y)$. Thus

$$\begin{aligned}
H_\alpha(X,Y) &= \frac{1}{1-\alpha}\log\left[\sum_{x,y}(p(x)p(y))^\alpha\right] \\
&= \frac{1}{1-\alpha}\log\left[\sum_{x,y} p(x)^\alpha p(y)^\alpha\right] \\
&= \frac{1}{1-\alpha}\log\left[\sum_x p(x)^\alpha \sum_y p(y)^\alpha\right] \\
&= \frac{1}{1-\alpha}\log\left[\sum_x p(x)^\alpha\right] + \frac{1}{1-\alpha}\log\left[\sum_y p(y)^\alpha\right]
\end{aligned}$$

$$= H_\alpha(X) + H_\alpha(Y).$$

(e) Consider the distribution $\mathbf{p} = [1/2, 1/4, 1/4]$. Then the Rényi entropy is

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log\left[(1/2)^\alpha + 2(1/4)^\alpha\right]$$

For example, $H_2(\mathbf{p}) = 1.415$.

On the other hand, if we evaluate the right-hand side of the grouping equation by combining the second and third probabilities, we get

$$\begin{aligned}
&H_\alpha(1/2, 1/4+1/4) + (1/4+1/4)H_\alpha\left(\frac{1/4}{1/4+1/4}, \frac{1/4}{1/4+1/4}\right) \\
&= H_\alpha(1/2, 1/2) + \frac{1}{2}H_\alpha(1/2, 1/2) \\
&= \frac{3}{2}H_\alpha(1/2, 1/2) \\
&\overset{(a)}{=} \frac{3}{2}\log 2 \\
&= \frac{3}{2}
\end{aligned}$$

where (a) follows from the result in part (a) of this problem, since $[1/2, 1/2]$ is a uniform distribution on $m = 2$ values. Since as computed above, $H_\alpha(\mathbf{p})$ is not always $3/2$, this example confirms that Rényi entropy does not satisfy the grouping rule.

7. Problem 2.32 from Cover-Thomas: *Fano.* We are given the following joint distribution on (X, Y):

		Y: a	b	c
X	1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $P_e = \Pr\{\hat{X}(Y) \neq X\}$.

(a) Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated probability P_e.

(b) Evaluate Fano's inequality for this problem and compare.

Solution:

(a) For a given function $\hat{X}(Y)$, the probability of error is given by

$$\begin{aligned}
P_e &= \Pr\{\hat{X}(Y) \neq X\} \\
&= \sum_{x,y:\hat{X}(y)\neq x} p(x,y) \\
&= \sum_y \sum_{x:x\neq\hat{X}(y)} p(x,y)
\end{aligned}$$

$$= \sum_y \Big(p(y) - p_{X,Y}(\hat{X}(y), y) \Big).$$

From this formula, it is clear that, to minimize the probability of error, for a given y, we should choose $\hat{X}(y)$ to maximize $p_{X,Y}(\hat{X}(y), y)$. Thus, the optimal choice is

$$\hat{X}(a) = 1, \quad \hat{X}(b) = 2, \quad \hat{X}(c) = 3.$$

The resulting probability of error is

$$\begin{aligned} P_e &= (p_Y(a) - p_{X,Y}(1,a)) + (p_Y(b) - p_{X,Y}(2,b)) + (p_Y(c) - p_{X,Y}(3,c)) \\ &= \left(\frac{1}{3} - \frac{1}{6}\right) + \left(\frac{1}{3} - \frac{1}{6}\right) + \left(\frac{1}{3} - \frac{1}{6}\right) \\ &= \frac{1}{2}. \end{aligned}$$

(b) Fano's inequality states that

$$H(Y|X) \le H(P_e) + P_e \log(|\mathcal{X}| - 1).$$

Given the calculation of P_e from the part (a), the right-hand side is

$$H(1/2) + \frac{1}{2}\log(3-1) = 1 + \frac{1}{2} = \frac{3}{2}.$$

To calculate $H(Y|X)$, note that for each x, the conditional distribution of Y given $X = x$ is some permutation of $[1/2, 1/4, 1/4]$. Thus

$$H(Y|X) = H(1/2, 1/4, 1/4) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = \frac{1}{2} + \frac{1}{4}(2) + \frac{1}{4}(2) = \frac{3}{2}.$$

Thus, in this case Fano's inequality holds with equality.

8. Problem 2.37 from Cover-Thomas: *Relative entropy.* Let X, Y, Z be three random variables with a joint probability mass function $p(x, y, z)$. The relative entropy between joint distribution and the product of the marginals is

$$D(p(x,y,z) \| p(x)p(y)p(z)) = E\left[\log \frac{p(x,y,z)}{p(x)p(y)p(z)}\right].$$

Expand this in terms of entropies. When is this quantity zero?

Solution:

$$\begin{aligned} D(p(x,y,z) \| p(x)p(y)p(z)) &= \mathbb{E}\left[\log \frac{p(x,y,z)}{p(x)p(y)p(z)}\right] \\ &= \mathbb{E}[\log p(x,y,z)] - \mathbb{E}[\log p(x)] - \mathbb{E}[\log p(y)] - \mathbb{E}[\log p(z)] \\ &= -H(X,Y,Z) + H(X) + H(Y) + H(Z). \end{aligned}$$

Furthermore, $D(p(x,y,z) \| p(x)p(y)p(z)) = 0$ if and only if $p(x,y,z) = p(x)p(y)p(z)$ for all (x,y,z), *i.e.* if X and Y and Z are independent.

9. Problem 2.42 from Cover-Thomas: *Inequalities.* Which of the following inequalities are generally $\geq, =, \leq$? Label each with $\geq$, $=$, or $\leq$.

   (a) $H(5X)$ vs. $H(X)$

   (b) $I(g(X);Y)$ vs. $I(X;Y)$

   (c) $H(X_0|X_{-1})$ vs. $H(X_0|X_{-1},X_1)$

   (d) $H(X,Y)/(H(X)+H(Y))$ vs. 1

   Solution:

   (a) $X \to 5X$ is a bijective mapping, and hence $H(X) = H(5X)$.

   (b) By the data processing inequality, $I(g(X);Y) \leq I(X;Y)$.

   (c) Because conditioning reduces entropy, $H(X_0|X_{-1}) \geq H(X_0|X_{-1},X_1)$.

   (d) $H(X,Y) \leq H(X)+H(Y)$, so $H(X,Y)/(H(X)+H(Y)) \leq 1$.