

# Homework #1

*Jonathan McFadden*

TCSS - 580 : Winter 2018

Information Theory

**Problem 2.3):** Let  $\mathbb{P}^n$  be the set of all  $n$ -dimensional probability vectors, with elements  $\vec{p} \in \mathbb{P}^n$  defined as  $\vec{p} = (p_1, p_2, \dots, p_i, \dots, p_n)$  for  $i \in \mathbb{Z}^+ \ni i \leq n$ . By the definition of a probability space, we must have

$$\vec{p} \cdot \vec{1} = \sum_{i=1}^n \{p_i\} = 1 \quad , \quad \forall \vec{p} \in \mathbb{P}^n, \quad (2.3-1)$$

where vector  $\vec{1}$  is defined as  $\vec{1} = (q_1, q_2, \dots, q_k, \dots, q_n) \in \mathbb{Z}^n$  with  $q_k = 1, \forall k \in [1, n]$  (where the interval  $[1, n]$  is defined such that  $[1, n] \subseteq \mathbb{Z}^+$ ). Furthermore, the definition of a probability space also requires that, for any  $\vec{p} \in \mathbb{P}^n$ , the elements of  $\vec{p}$  (the  $p_i \in \vec{p}$  such that  $i \in \mathbb{Z}^+ \ni i \leq n$ ) satisfy the condition

$$p_i \geq 0 \quad (2.3-2)$$

for all  $i \in \mathbb{Z}^+ \ni i \leq n$ .

The expression in 2.3-1 guarantees that the  $p_i$  of any  $\vec{p} \in \mathbb{P}^n$  satisfy the bound  $0 \leq p_i \leq 1$  where  $i \in \mathbb{Z}^+ \ni i \leq n$ . Therefore, the relation

$$p_i \log_2 [p_i] \geq 0 \quad (2.3-3)$$

holds for all  $p_i$  of any  $\vec{p} \in \mathbb{P}^n$ . Moreover, for the cases where  $p_i = 0$  or  $p_i = 1$ , it is clear that the expression in 2.3-3 reduces to equality. Specifically, the relation  $p_i \log_2 [p_i]$  becomes

$$p_i \log_2 [p_i] = 0 \quad (2.3-4)$$

for the case where  $p_i = 0$  or  $p_i = 1$ . Moreover, the relation in 2.3-4 also represents the **smallest** possible value/result for the expression  $p_i \log_2 [p_i]$ . That is to say, that when  $p_i = 0$  or  $p_i = 1$ , then  $p_i \log_2 [p_i]$  is at a minimum.

The result in 2.3-1, makes it is clear that *only* **ONE**  $p_i$  in each  $\vec{p} \in \mathbb{P}^n$  may have the value  $p_i = 1$ ; therefore the probability vectors  $\vec{p} \in \mathbb{P}^n$  which result in a minimum value for  $p_i \log_2 [p_i]$  all have exactly one non-zero element with the non-zero element having a value of one. This implies that there are only  $n$  such probability vectors,  $\vec{p}^*$  within any  $\mathbb{P}^n$ . Furthermore, the value of  $H(X) = \sum_{i=1}^n \{p_i \log_2 [p_i]\}$  for any such  $\vec{p}^*$  is also zero.

**Problem 2.4 a)**: Recall the chain-rule for conditional entropies of  $X$  given  $Y$ ,

$$H(X | Y) = H(X, Y) - H(Y) \quad (2.4-1)$$

We apply the expression in 2.4-1 to the case of  $g(X)$  given  $X$  to obtain

$$H(g(X) | X) = H(g(X), X) - H(X) \quad (0.0.1)$$

by rearranging the expression in the previous result as follows,

$$H(g(X), X) = H(X) + H(g(X) | X) \quad (2.4-2)$$

we obtain the desired result.

**Problem 2.4 b):** For any given value of  $X$ , we automatically know  $g(X)$ . Therefore, the expression for  $H(g(X), X)$  in 2.4-2 becomes

$$H(g(X), X) = H(X)$$

which is the desired result.

**Problem 2.4 c):** Recalling the expression for the conditional entropy chain rule in 2.4-1 and using it for the case where  $X = X$  and  $Y = g(X)$  yields the result

$$H(X | g(X)) = H(X, g(X)) - H(g(X))$$

rearranging the above expression yields

$$H(X, g(X)) = H(g(X)) + H(X | g(X)) \quad (2.4-3)$$

we obtain the desired result.

**Problem 2.4 d):** For any arbitrary function,  $g(X)$ , of a random variable  $X$ , the entropy  $H(X | g(X))$  satisfies the condition

$$H(X | g(X)) \geq 0 \quad (2.4-4)$$

for the case where  $g(X)$  is one-to-one, the relation in 2.4-4 simplifies to

$$H(X | g(X)) = 0$$

Applying the relation in 2.4-4 to the expression in 2.4-3 yields

$$\begin{aligned} H(X, g(X)) &= H(g(X)) + H(X | g(X)) \\ &\geq H(g(X)) + H(X | g(X)) - H(X | g(X)) \\ &\geq H(g(X)) \end{aligned}$$

**Problem 2.9 a):** Let  $\rho(X, Y)$  be a function which is defined according to

$$\rho(X, Y) = H(X | Y) + H(Y | X) \quad (2.9-1)$$

for all  $x$  and  $y$ . Since conditional probabilities are always non-zero (*for arbitrary  $X$  and  $Y$  we have  $H(X | Y) \geq 0$* ), we can say that  $H(X | Y)$  has the property

$$H(X | Y) \geq 0$$

and that  $H(Y | X)$  has the property

$$H(Y | X) \geq 0$$

Applying these properties of  $H(X | Y) \geq 0$  and  $H(Y | X) \geq 0$  to the expression in 2.9-1 yields

$$\rho(X, Y) = H(X | Y) + H(Y | X) \geq 0 \quad (2.9-2)$$

which indicates that  $\rho(X, Y)$  satisfies the first property of a metric over all  $x$  and  $y$ . By its definition in 2.9-1, we can say that  $\rho(X, Y)$  is symmetric; therefore, we can additionally say that  $\rho(X, Y)$  satisfies the second condition of a metric over all  $x$  and  $y$ .

Now, consider three random variables,  $X, Y$ , and  $Z$ . Then write

$$\rho(X, Y) = H(X | Y) + H(Y | X) \quad (2.9-3)$$

and

$$\rho(Y, Z) = H(Y | Z) + H(Y | Z) \quad (2.9-4)$$

and

$$\rho(X, Z) = H(X | Z) + H(Z | X) \quad (2.9-5)$$

We now add the expression in 2.9-3 and 2.9-4 to obtain

$$\begin{aligned} & H(X | Y) + H(Y | X) + H(Y | Z) + H(Z | Y) \\ & \left[ H(X | Y) + H(Y | Z) \right] + \left[ H(Z | Y) + H(Y | X) \right] \end{aligned} \quad (2.9-6)$$

By the chain rule for conditional entropies, we have  $H(X | Y) + H(Y | Z) = H(X, Y | Z)$  and  $H(Z | Y) + H(Y | X) = H(Z, Y | X)$  so the expression in 2.9-6 becomes

$$H(X, Y | Z) + H(Z, Y | X)$$

Again applying the chain rule for conditional entropies to the previous result, we have

$$\left[ H(X | Z) + H(Y | X, Z) \right] + \left[ H(Z | X) + H(Y | Z, X) \right]$$

Since conditional entropies are always greater or equal to zero, the  $H(Y | X, Z)$  and  $H(Y | Z, X)$  terms in the previous result satisfy  $H(Y | X, Z) \geq 0$  and  $H(Y | Z, X) \geq 0$ . This allows us to rewrite the previous result as

$$\begin{aligned}
 \rho(X, Y) + \rho(Y, Z) &= H(Y | Z) + H(Y | Z) + H(X | Y) + H(Y | X) \\
 &= \left[ H(X | Z) + H(Y | X, Z) \right] + \left[ H(Z | X) + H(Y | Z, X) \right] \\
 &\geq H(X | Z) + H(Z | X)
 \end{aligned}$$

Using the definition from 2.9-5, the previous result becomes

$$\begin{aligned}
 \rho(X, Y) + \rho(Y, Z) &= \left[ H(X | Z) + H(Y | X, Z) \right] + \left[ H(Z | X) + H(Y | Z, X) \right] \\
 &\geq H(X | Z) + H(Z | X) \\
 &\geq \rho(X, Z)
 \end{aligned}$$

thereby indicating that  $\rho(X, Y) + \rho(Y, Z) \geq \rho(X, Z)$  holds and, by extension, that the definition in 2.9-1 satisfies the fourth condition of a metric over all  $x$  and  $y$ .

Finally, we consider the case where  $X = Y$  via a one-to-one mapping. Since  $H(X, Y) = 0$  iff  $X$  is a function of  $Y$  and  $H(Y, X) = 0$  iff  $Y$  is a function of  $X$ ,  $\rho(X, Y)$  can only equal zero if and only if  $X = Y$ . This satisfies the third and only remaining condition of a metric over all  $x$  and  $y$ ; therefore, for cases where  $X$  and  $Y$  are related by a one-to-one mapping,  $\rho(X, Y)$  is a metric over all  $x$  and  $y$ .

**Problem 2.9 b)**: Starting with the expression  $I(X; Y) = H(X) - H(X | Y)$  we rearrange to obtain

$$H(X | Y) = H(X) - I(X; Y) \tag{2.9-7}$$

We also obtain

$$H(Y | X) = H(Y) - I(Y; X) \quad (2.9-8)$$

similarly. We now apply the expressions in 2.9-7 and 2.9-8 to the definition of  $\rho(X, Y)$  in 2.9-1 to obtain the result

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

Noting that  $I(X; Y) = I(Y; X)$  the previous result can be simplified to the expression

$$\begin{aligned} \rho(X, Y) &= H(X) - I(X; Y) + H(Y) - I(Y; X) \\ &= H(X) - I(X; Y) + H(Y) - I(X; Y) \\ &= H(X) + H(Y) - 2I(X; Y) \end{aligned} \quad (2.9-9)$$

which proves the first line of the problem. Next, we note that  $I(X; Y) = H(X) + H(Y) - H(X, Y)$  and apply this relation to the expression in 2.9-9 so that we obtain

$$\begin{aligned} \rho(X, Y) &= H(X) + H(Y) - 2I(X; Y) \\ &= H(X) + H(Y) - I(X; Y) - I(X; Y) \\ &= H(X) + H(Y) - I(X; Y) - [H(X) + H(Y) - H(X, Y)] \\ &= H(X, Y) - I(X; Y) \end{aligned} \quad (2.9-10)$$

as our result and proving the second line of the problem. Finally, we again note

$I(X; Y) = H(X) + H(Y) - H(X, Y)$  and then apply it to the expression in 2.9-10 to yield the result



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

which proves the third and final line of the problem.

**Problem 2.10 a):**

**Problem 2.10 b):**

Problem 2.11 a):

Problem 2.11 b):

Problem 2.11 c):

Problem 2.11 d):

**Problem 2.24 a):** Consider a choice of four unique objects and specify a specific, arbitrarily picked object to be "special". We can model this situation using two random variables  $X$  and  $Y$ . We define  $X$  to be the random variable describing whether the "special" object was picked, so  $\mathcal{X} = \{0, 1\}$  with  $X = 0$  if the "special" object was picked and  $X = 1$  otherwise. The entropy of this random variable,  $H(X)$ , is the entropy we seek to find, namely  $H(1/4)$ . Additionally, we define  $Y$  to be the random variable describing which of the three "non-special" items was picked, thus  $\mathcal{Y} = \{0, 1, 2\}$ . We define  $Y = 0$  if the first "non-special" object was picked,  $Y = 1$  for the second, and  $Y = 2$  for the third.

We can also define a third random variable  $Z$  which describes choosing one of four unique objects, thus  $\mathcal{Z} = \{0, 1, 2, 3\}$  and  $\Pr[Z = z] = 1/4$  for all  $z \in \mathcal{Z}$ . Furthermore, we can equate  $Z = XY$ , thus we can say

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

by also using the chain rule  $H(X, Y) = H(X) + H(Y | X)$ . Without loss of generality, we also say that the "special" object is represented by  $Z = 0$ . That is to say that  $H(X) = H(1/4) = H(Z = 0)$ .

We continue by applying the relation  $H(X) = H(1/4)$  to the previous result gives us

$$\begin{aligned} H(Z) &= H(X) + H(Y | X) \\ &= H(1/4) + H(Y | X) \end{aligned}$$

which becomes

$$\begin{aligned}
\longrightarrow H(1/4) &= H(Z) - H(Y | X) \\
&= H(Z) - \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \left\{ p(x, y) \log_2 [p(y|x)] \right\} \\
&= H(Z) - \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \left\{ \sum_z \{ p(z) p(y | x) \} \log_2 [p(y|x)] \right\} \\
&= H(Z) - \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \left\{ \sum_z \{ p(z) \} p(y | x) \log_2 [p(y|x)] \right\} \\
&= H(Z) - \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \left\{ \sum_z \{ p(z) \} p(y | x) \log_2 [p(y|x)] \right\} \\
&= H(Z) - \sum_{\substack{z \\ z \neq 0}} \left\{ \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \{ p(z) p(y | x) \log_2 [p(y|x)] \} \right\} \\
&= H(Z) - \sum_{\substack{z \\ z \neq 0}} \left\{ p(z) \left( \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \{ p(y | x) \log_2 [p(y|x)] \} \right) \right\} \tag{2.24-1}
\end{aligned}$$

through some simple rearranging along with noting the definition of  $H(Y | X)$  and the fact that  $Z = XY$  implies that  $p(z) = p(x, y)$ . Now, since  $\Pr[Z = z] = 1/4$  for all  $z \in \mathcal{Z}$  we can compute the value of  $H(Z)$  to be

$$H(Z) = \sum_{i=1}^4 \left\{ \frac{1}{4} \log_2 \left[ \frac{1}{4} \right] \right\} = 2 \tag{2.24-2}$$

Moreover, we can also evaluate the summation over  $z$  in the second term of 2.24-1 to obtain

$$\sum_{\substack{z \\ z \neq 0}} \left\{ p(z) \left( \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \{ p(y | x) \log_2 [p(y|x)] \} \right) \right\} = \frac{3}{4} \left( \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \{ p(y | x) \log_2 [p(y|x)] \} \right)$$

Additionally, we can say that, when  $X = 1$ ,  $p(y|x) = 1/3$  for all  $y \in \mathcal{Y}$ . Therefore, we can evaluate the

summation over  $x$  and  $y$  in the previous result

$$\sum_{\substack{z \\ z \neq 0}} \left\{ p(z) \left( \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \{ p(y | x) \log_2 [p(y|x)] \} \right) \right\} = \frac{3}{4} \sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \left\{ p(y | x) \log_2 [p(y|x)] \right\}$$

**Problem 2.24 b):**

**Problem 2.24 c):**

Problem 2.29 a):

Problem 2.29 b):

Problem 2.29 c):

Problem 2.29 d):



**Problem 2.31):**

**Problem 2.35):**

Problem 2.42 a):

Problem 2.42 b):

Problem 2.42 c):

Problem 2.42 d):