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 $\vec{p} = (p_1, p_2, \ldots, p_i, \ldots, p_n)$ for $i \in \mathbb{Z}^+ \ni i \leq n$. By the definition of a probability space, we must have

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

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

$$p_i \geq 0 \tag{2.3-2}$$

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

$$p_i \log_2 [p_i] \geq 0 \qquad (2.3\text{-}3)$$

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 realation $p_i \log_2 [p_i]$ becomes

$$p_i \log_2 [p_i] = 0 \qquad (2.3\text{-}4)$$

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 <u>only</u> **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, $\vec{p}^{\star}$ within any $\mathbb{P}^n$. Furthermore, the value of $H(X) = \sum_{i=1}^{n} \{p_i \log_2 [p_i]\}$ for any such $\vec{p}^{\star}$ is also zero.

<u>**Problem 2.4 a):**</u> Recall the chain-rule for conditional entropies of X given Y,

$$H\left(X \mid Y\right) = H\left(X, Y\right) - H\left(Y\right) \qquad (2.4\text{-}1)$$

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

$$H(g(X) \mid X) = H(g(X), X) - H(X) \tag{0.0.1}$$

by rearranging the expression in the previous result as follows,

$$H(g(X), X) = H(X) + H(g(X) \mid X) \tag{2.4-2}$$

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

$$H(X \mid g(X)) = H(X, g(X)) - H(g(X))$$

rearranging the above expression yields

$$H\left(X, g\left(X\right)\right) = H\left(g\left(X\right)\right) + H\left(X \mid g\left(X\right)\right) \tag{2.4-3}$$

we obtain the desired result.

Problem 2.4 d): For any arbitrary function, $g\left(X\right)$, of a random variable X, the entropy $H\left(X \mid g\left(X\right)\right)$

satisfies the condition

$$H\left(X \mid g\left(X\right)\right) \geq 0 \tag{2.4-4}$$

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

$$H\left(X \mid g\left(X\right)\right) = 0$$

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

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

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

$$\rho(X, Y) = H(X \mid Y) + H(Y \mid X) \tag{2.9-1}$$

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

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

and that $H(Y \mid X)$ has the property

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

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

$$\rho(X, Y) = H(X \mid Y) + H(Y \mid X) \geq 0 \tag{2.9-2}$$

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

$$\rho(X, Y) = H(X \mid Y) + H(Y \mid X) \tag{2.9-3}$$

and

$$\rho\left(Y, Z\right) = H\left(Y \mid Z\right) + H\left(Y \mid Z\right) \tag{2.9-4}$$

and

$$\rho\left(X, Z\right) = H\left(X \mid Z\right) + H\left(Z \mid X\right) \tag{2.9-5}$$

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

$$H\left(X \mid Y\right) + H\left(Y \mid X\right) + H\left(Y \mid Z\right) + H\left(Z \mid Y\right)$$

$$\left[H\left(X \mid Y\right) + H\left(Y \mid Z\right)\right] + \left[H\left(Z \mid Y\right) + H\left(Y \mid X\right)\right] \tag{2.9-6}$$

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

$$H\left(X, Y \mid Z\right) + H\left(Z, Y \mid X\right)$$

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

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

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

$$\rho\left(X,Y\right)+\rho\left(Y,Z\right)=H\left(Y\mid Z\right)+H\left(Y\mid Z\right)+H\left(X\mid Y\right)+H\left(Y\mid X\right)$$

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

$$\geq H\left(X\mid Z\right)+H\left(Z\mid X\right)$$

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

$$\rho\left(X,Y\right)+\rho\left(Y,Z\right)=\left[H\left(X\mid Z\right)+H\left(Y\mid X,Z\right)\right]+\left[H\left(Z\mid X\right)+H\left(Y\mid Z,X\right)\right]$$

$$\geq H\left(X\mid Z\right)+H\left(Z\mid X\right)$$

$$\geq\rho\left(X,Z\right)$$

thereby indicating that $\rho\left(X,Y\right)+\rho\left(Y,Z\right)\geq\rho\left(X,Z\right)$ 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 $H\left(X,Y\right)=0$ iff X is a function of Y and $H\left(Y,X\right)=0$ iff Y is a function of X, $\rho\left(X,Y\right)$ 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, $\rho\left(X,Y\right)$ is a metric over all x and y.

Problem 2.9 b): Starting with the expression $I\left(X;Y\right)=H\left(X\right)-H\left(X\mid Y\right)$ we rearrange to obtain

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

We also obtain

$$H\left(Y \mid X\right) = H\left(Y\right) - I\left(Y;X\right) \tag{2.9-8}$$

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

$$\rho\left(X,Y\right) = H\left(X \mid Y\right) + H\left(Y \mid X\right)$$
$$= H\left(X\right) - I\left(X;Y\right) + H\left(Y\right) - I\left(Y;X\right)$$

Noting that $I\left(X;Y\right) = I\left(Y;X\right)$ the previous result can be simplied to the expression

$$\rho\left(X,Y\right) = H\left(X\right) - I\left(X;Y\right) + H\left(Y\right) - I\left(Y;X\right)$$
$$= H\left(X\right) - I\left(X;Y\right) + H\left(Y\right) - I\left(X;Y\right)$$
$$= H\left(X\right) + H\left(Y\right) - 2\,I\left(X;Y\right) \tag{2.9-9}$$

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

$$\rho\left(X,Y\right) = H\left(X\right) + H\left(Y\right) - 2\,I\left(X;Y\right)$$
$$= H\left(X\right) + H\left(Y\right) - I\left(X;Y\right) - I\left(X;Y\right)$$
$$= H\left(X\right) + H\left(Y\right) - I\left(X;Y\right) - \left[H\left(X\right) + H\left(Y\right) - H\left(X,Y\right)\right]$$
$$= H\left(X,Y\right) - I\left(X;Y\right) \tag{2.9-10}$$

as our result and proving the second line of the problem. Finally, we again note $I\left(X;Y\right) = H\left(X\right) + H\left(Y\right) - H\left(X,Y\right)$ and then apply it to the expression in 2.9-10 to yield the result

$$\rho\left(X,Y\right) = H\left(X,Y\right) - I\left(X;Y\right)$$
$$= H\left(X,Y\right) - \left[H\left(X\right) + H\left(Y\right) - H\left(X,Y\right)\right]$$
$$= 2\,H\left(X,Y\right) - H\left(X\right) - H\left(Y\right)$$

which proves the third and final line of the problem.

<u>Problem 2.10 a):</u>

<u>Problem 2.10 b):</u>

<u>**Problem 2.11 a):**</u>

<u>**Problem 2.11 b):**</u>

<u>**Problem 2.11 c):**</u>

<u>**Problem 2.11 d):**</u>

<u>**Problem 2.24 a):**</u> 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

$$H(Z) = H(X, Y)$$

$$= H(X) + H(Y \mid X)$$

by also using the chain rule $H(X, Y) = H(X) + H(Y \mid X)$. 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 appling the relation $H(X) = H(1/4)$ to the previous result gives us

$$H(Z) = H(X) + H(Y \mid X)$$

$$= H(1/4) + H(Y \mid X)$$

which becomes

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

through some simple rearranging along with noting the definition of $H\left(Y\mid X\right)$ and the fact that $Z = XY$

implies that $p\left(z\right) = p\left(x,y\right)$. Now, since $\Pr\left[Z = z\right] = 1/4$ for all $z \in \mathcal{Z}$ we can compute the value of $H\left(Z\right)$

to be

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

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

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

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

summation over x and y in the previous result

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

$$= \frac{3}{4} \left(\cancel{3} \left(\frac{\cancel{1}}{\cancel{3}} \log_2 \left[\frac{1}{3}\right] \right) \right)$$

$$= \frac{3}{4} \log_2 \left[\frac{1}{3}\right]$$

$$= -\frac{3}{4} \log_2 \left[3\right] \qquad\qquad (2.24\text{-}3)$$

We can now apply the values from 2.24-2 and 2.24-3 to the expression in 2.24-1 to obtain

$$H\left(1/4\right) = H\left(Z\right) + \sum_{\substack{z \\ z \neq 0}} \left\{ p\left(z\right) \left(\sum_{\substack{x,y \\ x=1 \\ y \in \mathcal{Y}}} \left\{ p\left(y \mid x\right) \log_2 \left[p\left(y|x\right)\right] \right\} \right) \right\}$$

$$= 2 + \left(-\frac{3}{4} \log_2 \left[3\right] \right)$$

$$= \boxed{2 - \frac{3}{4} \log_2 \left[3\right] \approx 0.811278 \text{ bits}}$$

as our result and final answer.

<u>Problem 2.24 b):</u>

<u>Problem 2.24 c):</u>

<u>Problem 2.29 a)</u>:

<u>Problem 2.29 b)</u>:

<u>Problem 2.29 c)</u>:

<u>Problem 2.29 d)</u>:

Problem 2.31):

<u>Problem 2.35):</u>

Problem 2.42 a):

Problem 2.42 b):

Problem 2.42 c):

Problem 2.42 d):