<u>Homework #1</u>

<u>Jonathan McFadden</u>

TCSS - 580 : Winter 2018

Information Theory

<u>Problem 2.3</u>): 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

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

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

$$p_i \log_2 [p_i] \geq 0 \tag{2.3-3}$$

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 realation $p_i \log_2 [p_i]$ becomes

$$p_i \log_2 [p_i] = 0 \tag{2.3-4}$$

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 <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,

$$H(X \mid Y) = H(X, Y) - H(Y) \tag{2.4-1}$$

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

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

by rearranging the expression in the previous result as follows,

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

we obtain the desired result.

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

$$H\left(g\left(X\right), X\right) = H\left(X\right)$$

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 $Y = g\left(X\right)$ yields the result

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

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, $g(X)$, of a random variable X, the entropy $H(X \mid g(X))$

satisfies the condition

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

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

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

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

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

<u>**Problem 2.9 a):**</u> 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 (<u>for arbitrary</u> X <u>and</u> Y <u>we have</u> $H(X \mid Y) \geq 0$), 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

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

and

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

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

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

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

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

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

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

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

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

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

$$= 2H\left(X,Y\right) - H\left(X\right) - H\left(Y\right)$$

which proves the third and final line of the problem.

Problem 2.10 a): Let X_1 and X_2 be discrete R.V.s having PMFs $p_1\left(\cdot\right)$ and $p_2\left(\cdot\right)$, respectively. Furthermore, let the alphabets of X_1 and X_2 be denoted as $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. These alphabets are defined $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m+1, \ldots, n\}$, where $m, n \in \mathbb{Z}^+ nim < n$. Now define another R.V. X such that

$$X = \begin{cases} X_1, & \text{with probability} \quad \alpha \\ \\ X_2, & \text{with probability} \quad 1 - \alpha \end{cases}$$

In order to find $H\left(X\right)$ in terms of $H\left(X_1\right)$, $H\left(X_2\right)$, and α, we must first define the function, of X, $\theta = f\left(X\right)$, as follows

$$f\left(X\right) = \begin{cases} 1, & \text{when} \quad X = X_1 \\ \\ 2, & \text{when} \quad X = X_2 \end{cases}$$

Noting that $H\left(X\right) = H\left(X, f\left(X\right)\right) = H\left(f\left(X\right)\right) + H\left(X \mid f\left(X\right)\right)$, we can write

$$H(X) = H(X, f(X))$$

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

$$= H(f(X)) + f(f(X) = 1) H(X \mid f(X) = 1) + f(f(X) = 2) H(X \mid f(X) = 2)$$

$$= \boxed{H(\alpha) + \alpha H(X_1) + (1 - \alpha) H(X_2)} \tag{2.10-1}$$

which gives $H(X)$ in terms of $H(X_1)$, $H(X_2)$, and α, where $H(\alpha)$ is the binary entropy function written, in terms of α, as $H(\alpha) = -\alpha \log_2 [\alpha] - (1 - \alpha) \log_2 [1 - \alpha]$.

Problem 2.10 b): Since the function $H(\alpha) = -\alpha \log_2 [\alpha] - (1 - \alpha) \log_2 [1 - \alpha]$ is greater than or equal to zero for all $\alpha \in [0, 1]$, we may rewrite the expression in 2.10-1 as

$$H(X) = H(\alpha) + \alpha H(X_1) + (1 - \alpha) H(X_2)$$

$$H(X) - H(\alpha) = \alpha H(X_1) + (1 - \alpha) H(X_2)$$

$$H(X) \leq H(X_1) + H(X_2) \tag{2.10-2}$$

by also noting that $\alpha H(X_1) \leq H(X_1)$ and $(1 - \alpha) H(X_2) \leq H(X_2)$. Taking the expression in 2.10-2 to the power 2 gives

$$2^{H(X)} \leq 2^{H(X_1) + H(X_2)} \tag{0.0.2}$$

We can interpret the expression in 0.0.2 as indicating that, while X_1 and X_2 may have disjoint alphabets, it is possible that some combinations of elements from these alphabets may yield identical entropies.

Problem 2.11 a): Starting with

$$\rho = 1 - \frac{H\left(X_2 \mid X_1\right)}{H\left(X_1\right)} \tag{2.11-1}$$

We obtain a common denominator and note that, since X_1 and X_2 are identically distributed,

$H\left(X_1\right) = H\left(X_2\right)$ to yield the result

$$\begin{aligned}
\rho &= 1 - \frac{H\left(X_2 \mid X_1\right)}{H\left(X_1\right)} \\
&= \frac{H\left(X_1\right) - H\left(X_2 \mid X_1\right)}{H\left(X_1\right)} \\
&= \frac{H\left(X_2\right) - H\left(X_2 \mid X_1\right)}{H\left(X_1\right)}
\end{aligned}$$

Now, since $I\left(X_1; X_2\right) = H\left(X_1\right) - H\left(X_1 \mid X_2\right) = H\left(X_2\right) - H\left(X_2 \mid X_1\right)$, the previous result finally becomes

$$\begin{aligned}
\rho &= \frac{H\left(X_2\right) - H\left(X_2 \mid X_1\right)}{H\left(X_1\right)} \\
&= \frac{I\left(X_2; X_1\right)}{H\left(X_1\right)} = \boxed{\frac{I\left(X_1; X_2\right)}{H\left(X_1\right)}}
\end{aligned} \tag{2.11-2}$$

since $I\left(X_2; X_1\right) = I\left(X_1; X_2\right)$, by the definition of mutual information.

Problem 2.11 b): Note that, by the definitions of both entropy and conditional entropy, the expressions

$$0 \le H\left(X_2 \mid X_1\right) \le \left(H\left(X_2\right) = H\left(X_1\right)\right) \tag{2.11-3}$$

and

$$0 \le H\left(X_1 \mid X_2\right) \le \left(H\left(X_1\right) = H\left(X_2\right)\right) \tag{2.11-4}$$

are both valid. Dividing 2.11-3 by $H(X_1)$ and 2.11-4 by $H(X_2)$, we obtain

$$0 \leq \frac{H(X_2 \mid X_1)}{H(X_1)} \leq 1$$

and

$$0 \leq \frac{H(X_1 \mid X_2)}{H(X_2)} \leq 1$$

Since these quantities are bounded by $[0, 1]$, the expression in 2.11-1 then implies that $0 \leq \rho \leq 1$.

Problem 2.11 c): Recalling the result in 2.11-2, it is clear that the only way for $\rho = 0$ to hold, is for $I(X_1; X_2) = 0$ to also hold. Since $I(X_1; X_2) = 0$ only holds for independent X_1 and X_2, $\rho = 0$ can hold iff X_1 and X_2 are independent.

Problem 2.11 d): Again, from the expression in 2.11-1, it is clear that $\rho = 1$ only if $H(X_2 \mid X_1) = 0$. For this to be the case, X_2 __**must**__ be a function of X_1. Additionally, that function must be **one-to-one**, since the relation in 2.10-2 implies symmetry which means that X_1 can also be represented as a function of X_2.

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

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

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

$$\longrightarrow H\left(1/4\right) = H\left(Z\right) - H\left(Y \mid X\right)$$

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

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

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

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

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

$$= H\left(Z\right) + \sum_{\substack{z \\ z\neq 0}} \left\{ p\left(z\right) \left(\sum_{\substack{x,y \\ x=1 \\ y\in\mathcal{Y}}} \left\{ p\left(y \mid x\right) \log_2\left[p\left(y|x\right)\right] \right\} \right) \right\} \qquad (2.24\text{-}1)$$

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

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

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

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

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

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

$$= -\frac{3}{4} \log_2 \left[3\right] \tag{2.24-3}$$

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.

Problem 2.24 b): Let $f\left(x\right)$ be an arbitrary function which maps $f : x \in \mathbb{R} \to y \in \mathbb{R}$ where $y = f\left(x\right) \in \mathbb{R}$ for $x, y \in \mathbb{R}$. The average value of any function such as $f\left(x\right)$ over any interval covered by the function, $[a, b] \subseteq \mathbb{R} \ni a < b$, is defined according to the integral

$$\overline{f}\left(x \in [a, b]\right) = \frac{1}{b - a} \int_a^b f\left(x\right) \, dx$$

so that the average value can be determined by evaluating the integral expression.

When we consider the average entropy, $\overline{H}(p)$ of a system having uniform distribution, our bounds are clearly $[0, 1]$. Thus, since $H(p) = -p\log_2[p] - (1-p)\log_2[1-p]$, the above integral expression for the average value of a function becomes

$$\overline{H}(p \in [0,1]) = \frac{1}{\cancel{1}-0}\int_0^1 H(p)\ dp$$
$$= \int_0^1 \left(-p\log_2[p] - (1-p)\log_2[1-p]\right)dp \qquad (2.24\text{-}4)$$

Using Mathematica to evaluate the integral in 2.24-4, we obtain

$$\overline{H}(p \in [0,1]) = \frac{1}{\cancel{1}-0}\int_0^1 H(p)\ dp$$
$$= \int_0^1 \left(-p\log_2[p] - (1-p)\log_2[1-p]\right)dp$$
$$= \boxed{0.721}$$

Problem 2.29 a): Starting with $H(X,Y \mid Z)$, we apply the chain rule for entropies to obtain

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

Since all entropies are greater than or equal to zero (in this case $H(Y \mid X,Z)$), the previous result can be rewritten as

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

<u>**Problem 2.29 b):**</u> Starting with $I(X, Y; Z)$, we apply the chain rule for mutual information to obtain

$$I(X, Y; Z) = I(X; Z) + I(Y; Z \mid X)$$

Since mutual information is always greater than or equal to zero (<u>in this case</u> $I(Y; Z \mid X)$), the previous result becomes

$$I(X, Y; Z) = I(X; Z) + I(Y; Z \mid X)$$
$$\geq I(X; Z)$$

<u>**Problem 2.29 c):**</u> Starting with $H(X, Y, Z) - H(X, Y)$, and noting that $H(X, Y) = H(X) + H(Y \mid X)$ and $H(X, Y, Z) = H(X) + H(Y \mid X) + H(Z \mid X, Y)$, we have

$$H(X, Y, Z) - H(X, Y) = \cancel{H(X)} + \cancel{H(Y \mid X)} + H(Z \mid X, Y) - \left(\cancel{H(X)} + \cancel{H(Y \mid X)}\right)$$
$$= H(Z \mid X, Y) \tag{2.29-1}$$

By the definition of conditional mutual information, we have

$$I(Y; Z \mid X) = H(Z, X) - H(Z \mid X, Y)$$
$$\implies H(Z \mid X, Y) = H(Z, X) - I(Y; Z \mid X)$$

Applying the above relation to the expression in 2.29-1 gives

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

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

$$\geq H\left(Z,X\right) \qquad\qquad (2.29\text{-}2)$$

since mutual information is always greater than or equal to zero.

Problem 2.29 d): Noting that $I\left(X,Y;Z\right) = I\left(Z;X,Y\right)$ can be written as

$$I\left(X,Y;Z\right) = I\left(Z;X,Y\right)$$

$$I\left(X;Z \mid Y\right) + I\left(Z;Y\right) = I\left(Z;Y \mid X\right) + I\left(X;Z\right)$$

$$I\left(X;Z \mid Y\right) = I\left(Z;Y \mid X\right) + I\left(X;Z\right) - I\left(Z;Y\right)$$

thereby yeilding the desired result.

Problem 2.31): Starting with $H\left(X \mid g\left(Y\right)\right) = H\left(X \mid Y\right)$, we multiply both sides by -1 before adding $H\left(X\right)$ to both sides to obtain

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

$$H\left(X\right) - H\left(X \mid g\left(Y\right)\right) = H\left(X\right) - H\left(X \mid Y\right)$$

$$I\left(X;g\left(X\right)\right) = I\left(X;Y\right) \qquad\qquad (2.31\text{-}1)$$

by noting the definition of mutual information. The expression in 2.31-1 implies that in order for $H\left(X \mid g\left(Y\right)\right) = H\left(X \mid Y\right)$ to be true, $I\left(X;g\left(X\right)\right) = I\left(X;Y\right)$ must hold.

Problem 2.35): Using the definitions of entropy and mutual entropy, we have

$H(p) = 1.5$

$H(q) = 1.58$

$D(p \parallel q) = 0.084$

$D(q \parallel p) = 0.081$

Problem 2.42 a): Since the mapping is one-to-one $H(X) = H(5X)$

Problem 2.42 b): Since $X \longrightarrow g(X) \longrightarrow Y$ is a Markov Chain, the **Data Processing Inequality** assures us that $I(g(X); Y) \leq I(X; Y)$

Problem 2.42 c): Since conditioning can only reduce entropy, $H(X_0 \mid X_{-1}) \geq H(X_0 \mid X_{-1}, X_1)$.

Problem 2.42 d): Since Huffman coding is a one-to-one mapping, $H(X_1, X_2, \ldots, X_n) =$
$= H(c(X_1, X_2, \ldots, X_n))$

Problem 2.42 e): Since $H(X, Y) \leq H(X) + H(Y)$ we have

$$\frac{H(X, Y)}{H(X) + H(Y)} \leq 1$$