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 $q_k = 1, \forall k \in [1, n]$ (where the interval $[1, n]$ is defined such that $[1, n] \subseteq \mathbb{Z}^+$). Futhermore, the definition of a probability space also requires that, for any $\vec{p} \in \mathbb{P}^n$, the elements of $\vec{p}$ (the $p_i \in \vec{p}$ such that $i \in \mathbb{Z}^+ \ni i \leq n$) 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 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, $\vec{p}^\star$ 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.

Problem 2.4 a): 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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{gathered} H(X \mid Y) + H(Y \mid X) + H(Y \mid Z) + H(Z \mid Y) \\ \Big[H(X \mid Y) + H(Y \mid Z)\Big] + \Big[H(Z \mid Y) + H(Y \mid X)\Big] \end{gathered} \tag{2.9-6}$$

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

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

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

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

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

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

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 $I(X;Y) = H(X) - H(X \mid Y)$ we rearrange to obtain

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

We also obtain

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
H(X) &= H(X, f(X)) \\
&= H(f(X)) + H(X \mid f(X)) \\
&= H(f(X)) + f(f(X)=1)\, H(X \mid f(X)=1) + f(f(X)=2)\, H(X \mid f(X)=2) \\
&= \boxed{H(\alpha) + \alpha H(X_1) + (1-\alpha) H(X_2)} \qquad (2.10\text{-}1)
\end{aligned}
$$

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

$$
\begin{aligned}
H(X) &= H(\alpha) + \alpha H(X_1) + (1-\alpha) H(X_2) \\
H(X) - H(\alpha) &= \alpha H(X_1) + (1-\alpha) H(X_2) \\
H(X) &\leq H(X_1) + H(X_2) \qquad (2.10\text{-}2)
\end{aligned}
$$

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

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

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 both

$$0 \leq H\left(X_2 \mid X_1\right) \tag{0.0.3}$$

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 $Z = X\,Y$, thus we can say

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

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

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

which becomes

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

through some simple rearranging along with noting the definition of $H(Y \mid X)$ 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

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

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

$$\begin{aligned}\sum_{\substack{z \\ z\neq 0}}\left\{p(z)\left(\sum_{\substack{x,y \\ x=1 \\ y\in\mathcal{Y}}}\{p(y\mid x)\log_2[p(y|x)]\}\right)\right\} &= \frac{3}{4}\sum_{\substack{x,y \\ x=1 \\ y\in\mathcal{Y}}}\left\{p(y\mid x)\log_2[p(y|x)]\right\} \\ &= \frac{3}{4}\left(\cancel{3}\left(\frac{1}{\cancel{3}}\log_2\left[\frac{1}{3}\right]\right)\right) \\ &= \frac{3}{4}\log_2\left[\frac{1}{3}\right] \\ &= -\frac{3}{4}\log_2[3] \end{aligned} \tag{2.24-3}$$

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

$$\begin{aligned} H(1/4) = H(Z) + \sum_{\substack{z \\ z\neq 0}}\left\{p(z)\left(\sum_{\substack{x,y \\ x=1 \\ y\in\mathcal{Y}}}\{p(y\mid x)\log_2[p(y|x)]\}\right)\right\} \\ = 2 + \left(-\frac{3}{4}\log_2[3]\right) \\ = \boxed{2 - \frac{3}{4}\log_2[3] \approx 0.811278 \text{ bits}} \end{aligned}$$

as our result and final answer.

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

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

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

$$\begin{aligned}
\overline{H}(p \in [0,1]) &= \frac{1}{\cancel{1-0}} \int_0^1 H(p)\; dp \\
&= \int_0^1 \Big(-p\log_2[p] - (1-p)\log_2[1-p] \Big) dp
\end{aligned} \tag{2.24-4}$$

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

$$\begin{aligned}
\overline{H}(p \in [0,1]) &= \frac{1}{\cancel{1-0}} \int_0^1 H(p)\; dp \\
&= \int_0^1 \Big(-p\log_2[p] - (1-p)\log_2[1-p] \Big) dp \\
&= \boxed{0.721}
\end{aligned}$$

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):