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

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

by rearranging the expression in the previous result as follows,

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

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 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

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

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 $\rho(X,Y)$ 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

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

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

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

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

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

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.