Chapter 2 Problems

2.2. Find a general inequality for $H(X)$ vs $H(Y)$. let $y = g(x)$. $p(y) = \sum_{x:y=g(x)} p(x)$ Meaning sum over all x given that $y = g(x)$. It is pretty clear that if there is a 1-1 mapping then $p(x) = p(y)$.

Consider a set of x that maps to a single y. For this set $\sum_{x:y=g(x)} p(x)\ log\ p(x) \leq \sum_{x:y=g(x)} p(x)\ log\ p(y) = p(y)\ log\ p(y)$ since $p(y) \geq p(x)$. Now extending the argument to the entire range of X and Y we obtain:

$$\begin{aligned} p(x) &\leq \sum_{x:y=g(x)} p(x) = p(y) \\ H(X) &= -\sum_x p(x)\ log\ p(x) \\ &= -\sum_y \sum_{x:y=g(x)} p(x)\ log\ p(x) \\ &\geq -\sum_y p(y)\ log\ p(y) \\ &= H(Y) \end{aligned}$$

a) $H(X)$ vs $H(Y = 2^X)$
Y and X are 1-1 so we have the special case of equality.

b) $H(X)$ vs $H(Y = cos(X))$
There are multiple values of X that could map to single values of Y with this function. Thus we can only guarentee the general case where $H(X) \geq H(Y)$

2.4. Show that the entropy of a function of X is $\leq$ than the entropy of X by justifying the following steps.

a) $H(X, g(X)) = H(X) + H(g(X)|X)$
This is justified by the chain rule.

b) $= H(X)$
$H(g(X)|X) = 0$ because $g(X)$ is known for any value of X. Thus $H(g(X)|X) = \sum_x p(x)\ H(g(X)|X = x) = \sum_x 0 = 0$ leaving $H(X)$

c) $H(X, g(X)) = H(g(X)) + H(X|g(X))$
Just a different application of the chain rule.

d) $\geq H(g(X))$
$H(X|g(X)) \geq 0$ and is equal to 0 when $g(X)$ gives an unambiguous mapping to one X. Otherwise it will be greater than 0. Thus $H(g(X)) \leq H(X, g(X))$

2.10. We have a set

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

a) What is $H(X)$ in term of $H(X_1), H(X_2), \alpha$?

$$\begin{aligned} \theta = f(X) &= \begin{cases} 1 & \text{when } X = X_1 \\ 2 & \text{when } X = X_2 \end{cases} \\ H(X) &= H(X, f(X)) = H(\theta) + H(X|\theta) \\ &= H(\theta) + p(\theta = 1)\ H(X|\theta = 1) + p(\theta = 2)\ H(X|\theta = 2) \\ &= H(\alpha) + \alpha\ H(X_1) + (1-\alpha)\ H(X_2) \end{aligned}$$

where $H(\alpha) = -\alpha\ log\ \alpha - (1-\alpha)\ log\ (1-\alpha)$

2.12. We are given that $p(x, y) = \left[\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3}\right]$ Find the following:

a) H(X), H(Y)

$$\begin{aligned} H(X) &= H(\frac{2}{3}, \frac{1}{3}) \\ &= \frac{2}{3}log(\frac{3}{2}) + \frac{1}{3}log(3) \\ &\approx 0.9183bits \\ H(Y) &= H(\frac{2}{3}, \frac{1}{3}) \\ &\approx 0.9183bits \end{aligned}$$

b) H(X—Y), H(Y—X)

$$\begin{aligned} H(X|Y) &= \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) \\ &= \frac{1}{3}H(\frac{1}{2}, \frac{1}{2}) + \frac{2}{3}H(0,1) \\ &= \frac{1}{3}1 + \frac{2}{3}0 \\ &= \frac{1}{3}bits \\ H(Y|X) &= \frac{2}{3}H(Y|X=0) + \frac{1}{3}H(Y|X=1) \\ &= \frac{2}{3}H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{3}H(0,1) \\ &= \frac{2}{3}bits \end{aligned}$$

c) H(X,Y)

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X) \\ &\approx 1.585bits \end{aligned}$$

d) H(Y) - H(Y—X)

$$\begin{aligned} &H(Y) - H(Y|X) \\ &\approx 0.251bits \end{aligned}$$

e) I(X;Y)

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &\approx 0.251bits \end{aligned}$$

f) See Figure 1 which shows the above equations on a Venn diagram.

2.15. Markov chains: $X_1 \to X_2 \to \cdots \to X_n$, have joint probabilities of the following form:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})$$

This means that for mutual information we can do some nice reduction.

$$\begin{aligned} I(X_1; X_2, \ldots, X_n) &= \\ let \;\; X_3, \ldots, X_n &= Y \\ I(X_1; X_2, Y) &= I(X_1; X_2) + I(X_1; Y|X_2) \end{aligned}$$

Now since this is a markov chain, and we know that X_1 and Y are completely independent we can say that $I(X_1; Y|X_2) = 0$ since $Y|X_2$ is just a subset of Y which as noted previously is entirely independent of X_1. Thus the whole equation simplifies to $I(X_1; X_2)$

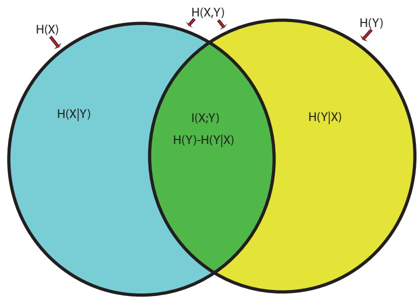

Figure 1: **2.12 Part e Venn Diagram**

2.26. a) $ln(x) \leq x - 1$ for $0 \leq x \leq \infty$ As x approaches 0 $x - 1$ approaches -1 while $ln(x)$ approaches $-\infty$. $x - 1$ grows linearly while $ln(x)$ grows much slower, logarithmically, so for large numbers it is given that $x - 1$ is larger.
From the derivative of $x-1$ and $ln(x)$ we know that $x-1$ has a slope of 1 for all x, while $ln(x)$ has slope $\frac{1}{x}$. This means that when $x < 1, ln(x)$ has an increasingly steep negative sloap, more steeply negative than 1, and when $x > 1$ it has an increasingly shallow positive slope. The only point at which their slopes are equal is $x = 1$, and since $ln(1) = 0 = 1-1$ we can say that $(x-1) \geq ln(x)$.

b) Since $D(p||q) = \sum_x p(x)ln\left(\frac{p(x)}{q(x)}\right)$ that means that $-D(p||q) = -\sum_x p(x)ln\left(\frac{p(x)}{q(x)}\right)$ which is equal to the reciprocal of the log, giving us: $\sum_x p(x)ln\left(\frac{q(x)}{p(x)}\right)$
Let $\frac{q(x)}{p(x)}$ be x, as shown before $x - 1 \geq ln(x)$ which justifies the second step.
The final step is fairly straight forward as well. In the equation $\sum_x p(x)\left(\frac{q(x)}{p(x)} - 1\right)$ the $p(x)$ terms cancel leaving $\sum_x q(x) - 1$. Since $q(x)$ is a probability, we know that over all x it will sum to 1 meaning that the equation simplifies to 0.

c) To equal 0, I have just shown that $\sum_x p(x)\left(\frac{q(x)}{p(x)} - 1\right) = 0$ additionally, in my argument for part a, I showed that $ln(x) = x - 1$ only when $x = 1$. Thus when $\frac{q(x)}{p(x)} = 1$ (when $q(x) = p(x)$) the value of D will be 0.

2.32. See attached sheet.