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 $H(X_1), H(X_2), \alpha$?

$$\theta = f(X) = \begin{cases} 1 & \text{when } X = X_1 \\ 2 & \text{when } X = X_2 \end{cases}$$

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

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.

$$\begin{aligned} &I(X_1; X_2, \ldots, X_n) = \\ &let \quad X_3, \ldots, X_n = Y \\ I(X_1; X_2, Y) = I(X_1; X_2) + I(X_1; Y|X_2) &\text{ partial application of chain rule} \end{aligned}$$

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 $I(X_1; X_2)$

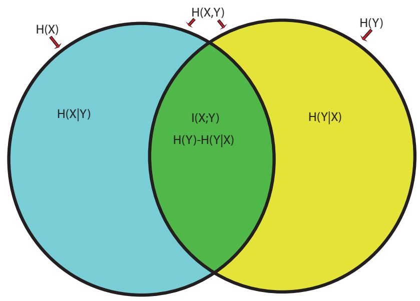

Figure 1: **2.12 Part e Venn Diagram**

2.26. a) $ln(x) \leq x - 1$ for $0 \leq x \leq \infty$ If the second derivative of a function is positive, that means that it is concave up, or convex. If it is always greater than 0, it will never not be convex. If it could be zero on the interval, then we can't make this claim, and the converse isn't necessarily true. Let $f(x) = x - 1 - ln\ x$ then

$$f'(x) = 1 - \frac{1}{x}$$
$$f''(x) = \frac{1}{x^2} > 0$$

so we can say that $f(x)$ is strictly convex and a local minimum is also a global minimum. To find the local minimum we just set $f'(x) = 0$ and get $x = 1$. So $f(x) \geq f(1)$ which means that $x - 1 - ln\ x \geq 1 - 1 - ln\ 1 = 0$. Showing that $x - 1 \geq ln\ x$ on the interval $(0, \infty)$.

b) Justify the following steps Let A be the set of x such that $p(x) > 0$

$$\begin{aligned} -D_e(p||q) &= \sum_{x \in A} p(x) ln \frac{q(x)}{p(x)} \\ &\leq \sum_{x \in A} p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \\ &= \sum_{x \in A} q(x) - \sum_{x \in A} p(x) \\ &\leq 0 \end{aligned}$$

The first step follows from the definition of D, the second step follows from the inequality $ln\ t \leq t - 1$, the third from expanding the sum, and the last step from the fact that the $q(A) \leq 1$ and $p(A) = 1$.

c) Conditions for equality? We have the inequality $ln\ t \leq t - 1$ from a previous problem, and showed equality iff $t = 1$. Therefore we have equality if $\frac{q(x)}{p(x)} = 1$ for all $x \in A$. This implies that $p(x) = q(x)$ for all x and then we'll have equality in the last step as well.

2.32.

$$P(X,Y) = \begin{pmatrix} \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{6} \end{pmatrix}$$

where X is indexed $[a, b, c]$ and Y is indexed $[1, 2, 3]$. Let $\hat{X}(Y)$ be an estimator for X based on Y and let $P_e = Pr\{\hat{X}(Y) \neq X\}$

a) From inspection we see that

$$\hat{X}(y) = \begin{cases} 1 & y = a \\ 2 & y = b \\ 3 & y = c \end{cases}$$

and the associated P_e is the sum of $P(1,b),\ P(1,c),\ P(2,a),\ P(2,c),\ P(3,a),\ P(3,b)$. Therefore $P_e = \frac{1}{2}$

b) From Fano's inequality we know

$$P_e \geq \frac{H(X|Y) - 1}{log\ |\chi|}$$

$$\begin{aligned}
H(X|Y) &= H(X|Y=a)Pr\{y=a\} + H(X|Y=b)Pr\{y=b\} + H(X|Y=c)Pr\{y=c\} \\
&= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) Pr\{y=a\} + H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) Pr\{y=b\} + H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) Pr\{y=c\} \\
&= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) (Pr\{y=a\} + Pr\{y=b\} + Pr\{y=c\}) \\
&= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) \\
&= 1.5 \text{ bits}
\end{aligned}$$

Our estimator of error $P_e \geq \frac{1.5-1}{log\ 3} = 0.316$ is not very close to Fano's bound in this form. Since $\hat{X} \in \chi$ we can use the stronger form of Fano's inequality $P_e \geq \frac{H(X|Y)-1}{log(|\chi|-1)}$ to get $P_e \geq \frac{1.5-1}{log\ 2} = \frac{1}{2}$ which is a great bound on our actual P_e.