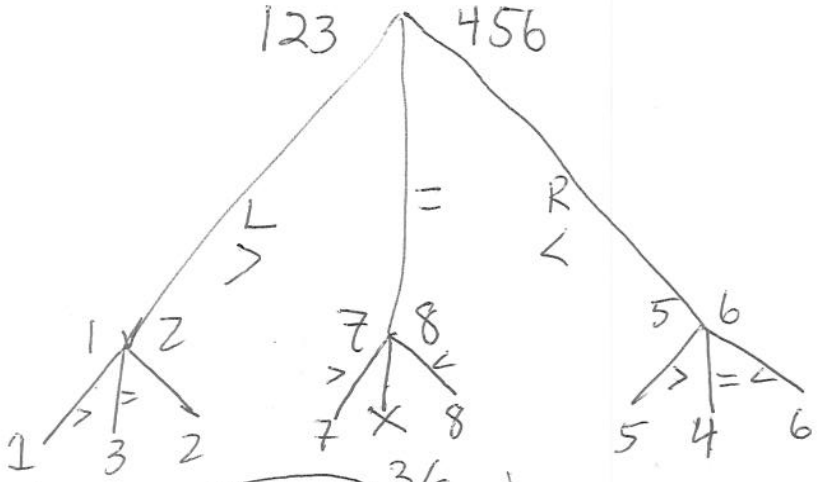
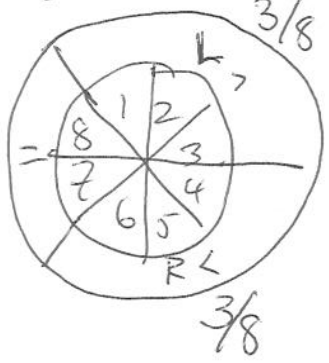
(3) Say you have 8 coins and one of them is heavier than the other 7, but you can't tell by looking at it or just by touch. Say that you do however have a balance scale which if you weigh coins against each other and will tell you if the left pan is heavy, right pan is heavy or both pans are equal.
   (a) If the location of the heavier coin is a random event, how much information is learned on average when you are told which coin is heavier?
   (b) If you really do gain $log_2(3) \approx 1.58$ bits of information per weighing, what is the maximum number of weighings that you should need to determine which of the 8 coins is the heavy one? Explain why (write a sentence!).
   (c) Draw a decision tree which demonstrates that it is possible to determine which of the coins is heavy in this number of weighings.
   (d) Say that you weigh coins 1 and 2 vs. 3 and 4 on your first weighing. How much information do you learn on average from that particular weighing?



$$H\left(\right.$$ 

$$\left.\right) = \frac{1}{4}\log_2 4 + \frac{3}{8}\log_2\left(\frac{8}{3}\right) \times 2$$

$$\frac{1}{2} + \frac{3}{4}\log_2\left(\frac{8}{3}\right) = 1.56$$

$$\approx$$

$$H(X|Y,Z) = \sum_{(Y,Z)=(a,b)} P(Y=a, Z=b)\, H(X|Y=a, Z=b) = 0$$

OR $Y=1$ & $Z=1$	$X=0$	$H(X\|Y=1 \& Z=1) = 0$
$Y=1$ & $Z=3$		
$Y=1$ & $Z=2$	$X=3$	$H(X\|Y=1 \& Z=2) = 0$
$Y=0$ & $Z=1,2,3$	$X=1$	$H(X\|Y=1 \& Z=1) = 0$
$Y=-1$ & $Z=1$	$X=0$	$H(X\|Y=-1, Z=1) = 0$
$Y=-1$ & $Z=2$ or 3	$X=-1$	$H(X\|Y=-1, Z=2) = 0$

$$H(Z|Y) = \sum_{Y=0,1,-1} P(Y=a)\, H(Z|Y=a)$$

$Y=0$ $Z=$ 

$H(Z|Y=0) = \log_2 3$

$Y=1$ or -1 $Z=$ 

$H(Z|Y=1) = \log_2 3$

$$H(Z|Y) = \overset{Y=0}{\frac{1}{4}}\log_2 3 + \overset{Y=1}{\frac{3}{8}}\log_2 3 + \overset{Y=-1}{\frac{3}{8}}\log_2 3 = \log_2 3$$

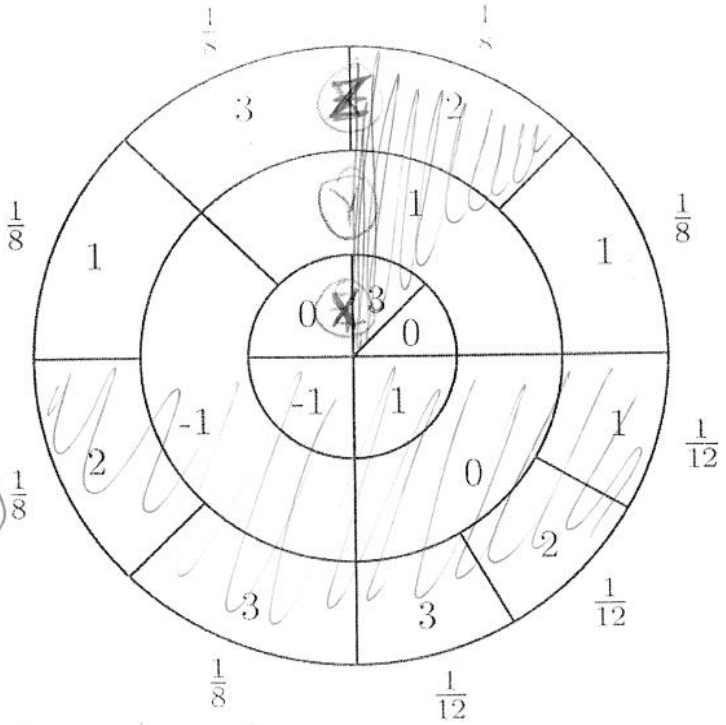
(a) Calculate $H[X]$.

(b) Calculate the expected number of binary registers needed to store Z.

(c) Calculate the uncertainty of Z given that $X = 0$.

(d) Calculate $H[X|Y, Z]$.

(e) Calculate $H[Z|Y]$.



$H(X) = \frac{1}{3}\log_2 3 + \frac{1}{3}\log_2 3 + \frac{1}{3}\log_2 3 = \log_2 3$

$H(Z) = \frac{3}{8}\log_2(8/3) + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8$

Information Theory Definitions

Definition: The conditional entropy of a random variable X given an event E

$$H(X \mid E) = \sum_a P[X = a \mid E] \; log_2 \left(\frac{1}{P[X = a \mid E]} \right)$$

Definition: The conditional entropy of X given Y

means "the amount of information that I learn from X given that I know Y

$$H(X \mid Y) = \sum_b P[Y = b] \; H(X \mid Y = b)$$

If X & Y are independent $H(X/Y) = H(X)$

If X is dependent on Y $H(X/Y) = 0$

Information Theory Definitions

Definition: The Entropy of a random variable X

$$H(X) = \sum_a P[X = a] \; log_2 \left(\frac{1}{P[X = a]} \right)$$

Expected value of the entropy of X

if X & Y are random variables then

X+Y means add the outcomes of the values of X & Y (only makes sense if X & Y take on numerical values).

Definition: The entropy of two random variables X and Y.

$$H(X, Y) = \sum_{a,b} P[X = a \;\&\; Y = b] \; log_2 \left(\frac{1}{P[X = a \;\&\; Y = b]} \right)$$

if X & Y are independent $H(X,Y) = H(X) + H(Y)$

if X is dependent on Y then $H(X,Y) = H(Y)$

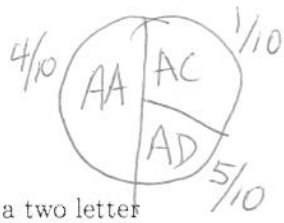
4. You cast a pair of dice: die #1 shows X and #2 shows Y but you only record $X + Y$.
   a) How much information have you lost in doing this?
   b) What is the expected number of bits you need to store 300 samples of $X + Y$?

6. Verbally explain why your intuition tells you that we must have H(Y|X)<=H(X) and when should we have equality.

(2) Random words are created by choosing the first letter using the following table of single letter statistics then each subsequent letter is chosen using the table of biletter statistics shown to the right.

A	4
B	1
C	2
D	3

	A	B	C	D
A	4	0	1	5
B	3	1	0	6
C	7	3	0	0
D	1	0	9	0



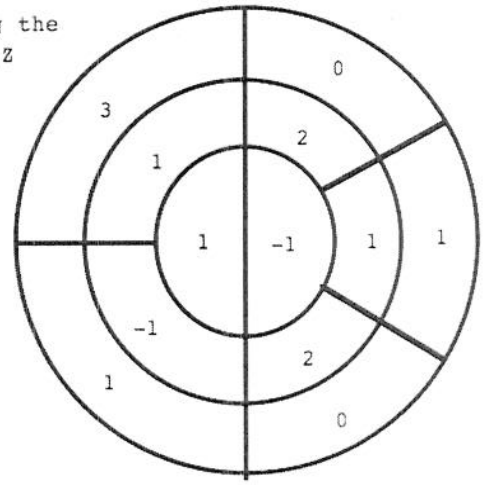
(a) How much information do you learn on average when you are told a two letter word given that you know that the word begins with the letter A?
(b) How much information do you learn on average when you are told a two letter word given that you know that the word ends with the letter D?
(c) How much information do you learn on average when you are told a two letter word given that you know at least one of the letters is a D?

3. Let X,Y,Z be produced by spinning the attached fortune wheel, with X,Y,Z given by the inner, middle and outer circles respectively. Calculate:

   a) H(Y|X,Z)

   b) H(X x Z)

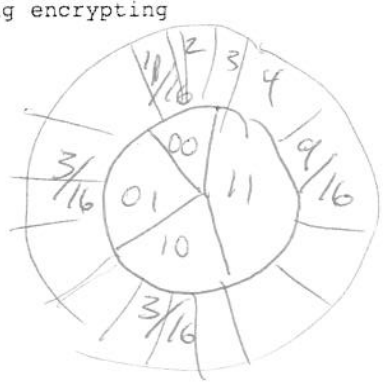
   c) The expected amount of information about Z you get when you are told the values of X and Y.



4. Suppose a random cryptosystem is obtained as follows: The message consists of two independent random variables X1 , X2 both generated by spinning a wheel with arcs labelled 0,1 of lengths 1/4 and 3/4 respectively. Let there be four equiprobable keys with the following corresponding encrypting transformations:

   k1: (X1,X2) ---> (X1,X2)
   k2: (X1,X2) ---> (X2,X1)
   k3: (X1,X2) ---> (X1, (X1+X2) mod 2)
   k4: (X1,X2) ---> ((X1+X2) mod 2 , X1)

   Calculate the entropy of the cyphertext.



If X is the roll of a die then $H(X) = \log_2 6$

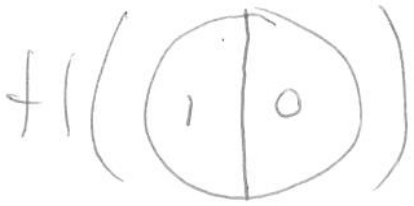
$H(X) + H(Y) = 2 \cdot \log_2 6 = 5.1669\ldots25$

$H(X+Y) = \frac{1}{36}\log_2 36 + \frac{1}{18}\log_2 18 + \frac{1}{12}\log_2 12 + \frac{1}{9}\log_2 9 + \frac{5}{36}\log_2 \frac{36}{5} + 1/6 \log_2 6$

$= 3.2744\ldots$

difference $= 1.8955\ldots$

minimum # of bits to store 300 X+Y's $= 300 \cdot 3.2744\ldots$
round up

$H\left(\right.$ 

$\left.\right) = 1$

$H\left(\right.$ 

$\left.\right) = \frac{1}{128}\log_2 128 + \frac{127}{128}\log_2\left(\frac{1}{1}\right)$

$= \frac{7}{128} + \frac{127}{128} \cdot \log_2\left(\frac{128}{127}\right)$

$128 = 2^7 \quad \log_2 128 = 7 \quad \approx .066$

Theorem 1 $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

Proof. Notice

$$P[X=a, Y=b] = P[X=a] \times \frac{P[X=a, Y=b]}{P[X=a]} = P[X=a] \times P[Y=b|X=a]$$

we may rewrite the definition of $H(X,Y)$ as

$$\begin{aligned}
H(X,Y) &= \sum_a \sum_b P[X=a, Y=b] \log_2 \frac{1}{P[X=a, Y=b]} \\
&= \sum_a \sum_b P[X=a, Y=b] \log_2 \frac{1}{P[X=a]P[Y=b|X=a]} \\
&= \sum_a \sum_b P[X=a, Y=b] \log_2 \frac{1}{P[X=a]} + \sum_a \sum_b P[X=a, Y=b] \log_2 \frac{1}{P[Y=b|X=a]} \\
&= \sum_a P[X=a] \log_2 \frac{1}{P[X=a]} + \sum_a \sum_b P[X=a]P[Y=b|X=a] \log_2 \frac{1}{P[Y=b|X=a]} \\
&= H(X) + \sum_a P[X=a] \sum_b P[Y=b|X=a] \log_2 \frac{1}{P[Y=b|X=a]} \\
&= H(X) + H(Y|X)
\end{aligned}$$

QED