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

$$H(X \mid Y) = \sum_{b} P[Y = b] \ H(X \mid Y = b)$$

amount of information that I learn when told X given that I know Y.

$$0 \leq H(X|Y) \leq H(X)$$

if X is dep on Y   if Y is indep of X

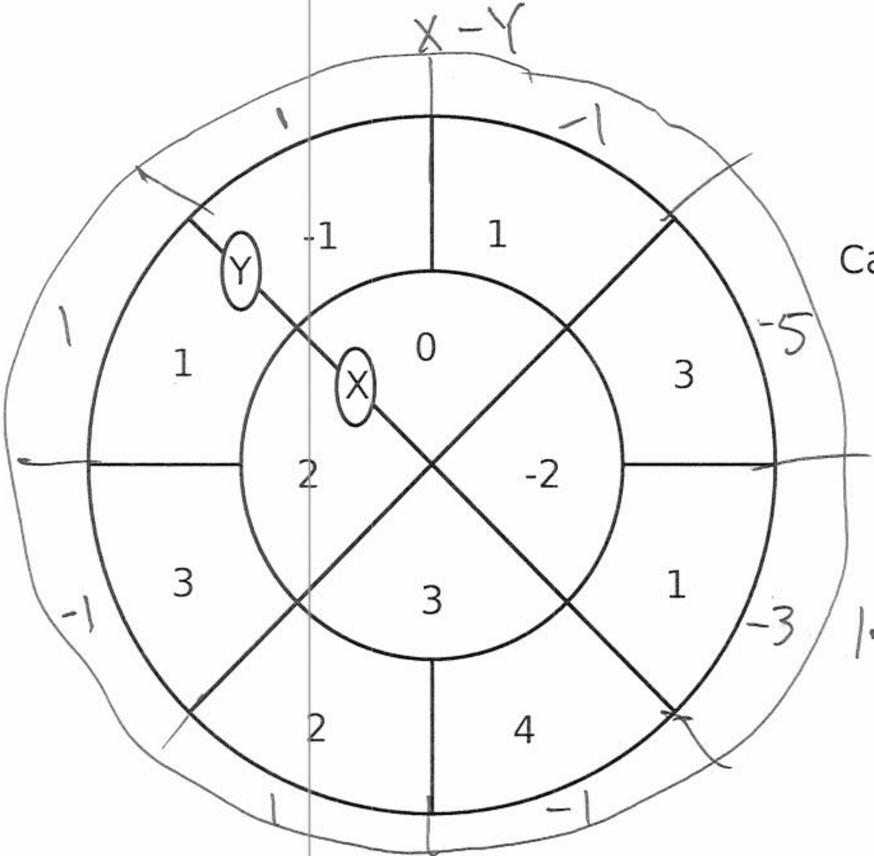
Information Theory Definitions

Definition: The Entropy of a random variable X

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

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

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



Calculate

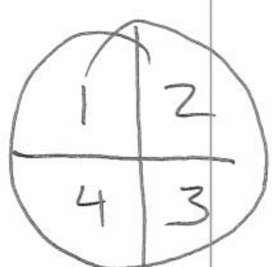
a) E(X) $= 3/4$

b) E(Y | X >0) $= 2.5$
$= 5/2$

c) E(X-Y) $=$

$$1\cdot\frac{3}{8} + (-1)\cdot\frac{3}{8} + (-3)\frac{1}{8} + (-5)\frac{1}{8}$$
$$= -1$$

$$E(X) = \cancel{\frac{1}{4}\cdot 0} + \cancel{\frac{1}{4}\cdot 2} + \cancel{\frac{1}{4}\cdot(-2)} + \frac{1}{4}\cdot 3 = 3/4$$



$Y|X>0$ $\quad E(Y|X>0) = \frac{1}{4}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 3 + 1/4 \cdot 4$

$$E(Y|X>0) = \sum_i i\cdot P(Y=i\,|\,\underline{\underline{X>0}})$$
$$= 1\cdot P(Y=1|X>0) + 2\cdot P(Y=2|X>0) + 3\cdot P(Y=3|X>0) + 4\cdot P(Y=4|X>0)$$
$$= 1\cdot\frac{P(Y=1\,\&\,X>0)}{P(X>0)} + 2\cdot\frac{P(Y=2\,\&\,X>0)}{P(X>0)} + 3\cdot\frac{P(Y=3\,\&\,X>0)}{P(X>0)} + 4\cdot\frac{P(Y=4\,\&\,X>0)}{P(X>0)}$$

Theorem 2 *For any two random variables X and Y we always have*

$$H(X|Y) \leq H(X) \tag{1}$$

and equality holds if and only if X and Y are independent.

Proof. From our definitions we get

$$\begin{aligned}
H(X|Y) &= \sum_b P[Y=b] H(X|Y=b) \\
&= \sum_b P[Y=b] \sum_a P[X=a|Y=b] \log_2 \frac{1}{P[X=a|Y=b]} \\
&= \sum_b \cancel{P[Y=b]} \sum_a \frac{P[X=a,Y=b]}{\cancel{P[Y=b]}} \log_2 \frac{1}{P[X=a|Y=b]} \\
P(X=a)\cdot P(Y=b|X=a) \quad &= \sum_b \sum_a \boxed{P[X=a,Y=b]} \log_2 \frac{1}{P[X=a|Y=b]} \\
&= \sum_a P[X=a] \boxed{\sum_b P[Y=b|X=a] \log_2 \frac{1}{P[X=a|Y=b]}} \qquad (2)
\end{aligned}$$

$$\leq \sum_a P(X=a) \log_2 \frac{1}{P(X=a)} =: H(X)$$

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

Proof. Notice

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

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

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

QED

Basic Identities and Inequalities

1. For any two random variables X and Y

$$\boxed{H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)}$$

2. For a random variable X which takes k distinct values

$$\boxed{H(X) \leq \log_2 k}$$

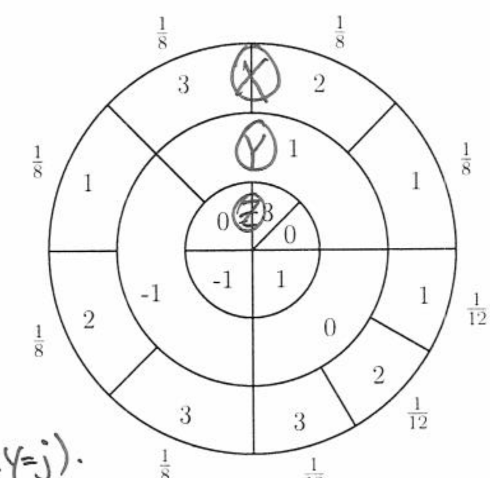
3. For a partition $A = \{A_1, A_2, \ldots, A_k\}$

$$\boxed{H(A) \leq \log_2 k}$$

4. For any two random variables X and Y

$$\boxed{\left.\begin{array}{l} H(X|Y) \leq H(X) \\ H(X,Y) \leq H(X) + H(Y) \end{array}\right\} \begin{array}{l}\text{equality if and only if } X \text{ and } Y \\ \text{are independent}\end{array}}$$

$$\boxed{H(X|Y) = 0 \Leftrightarrow X \text{ is a function of } Y}$$



(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 $H[Z|Y]$.

f) Calculate $H(Z|X \& Y) = 0$

$$= \sum_{i=\{1,2,3\}} \sum_{j=\{0,1,-1\}} P(X=i \,\&\, Y=j) \sum_{k=\{0,1,-1,3\}} P(Z=k|X=i \,\&\, Y=j) \log_2\left(\frac{1}{P(Z=k|X=i \,\&\, Y=j)}\right)$$

d) $H(X|Y;Z) = P(Y=1 \,\&\, Z=0)\,H(X|Y=1 \,\&\, Z=0) + P(Y=1 \,\&\, Z=3)\,H(X|Y=1 \,\&\, Z=3)$

$\frac{2}{8} \cdot 1 + \frac{1}{8} \cdot 0$

$+ P(Y=-1 \,\&\, Z=0)\,H(X|Y=-1 \,\&\, Z=0) + P(Y=-1 \,\&\, Z=-1)\,H(X|Y=-1 \,\&\, Z=-1)$

$+ \frac{1}{8} \cdot 0 + \frac{2}{8} \cdot 1$

$+ P(Y=0 \,\&\, Z=1) \cdot H(X|Y=0 \,\&\, Z=1)$

$+ \frac{1}{4} \cdot \log_2 3$

$= \frac{1}{2} + \frac{1}{4}\log_2 3$

$H(1\ 2\ 3) = \log_2 3$

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

$Y=1 \qquad Y=-1 \qquad Y=0$

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

Theorem 4 *For a random variable X which takes only k values we always have*

$$H(X) \leq \log_2 k$$

with equality if and only if X takes all its values with equal probability

Proof. The definition gives

$$H(X) = \sum_{b \in \text{VALUES}} P[X=b] \log_2 \frac{1}{P[X=b]}$$

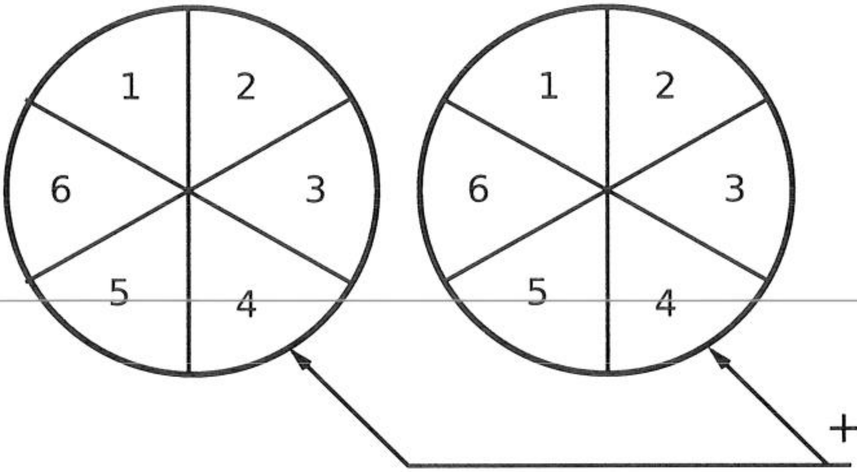
Using again the convex function inequality

$$\sum_b m_b \log_2 x_b \leq \log_2 \left(\sum_b m_b x_b \right)$$

gives

$$H(X) \leq \log_2 \left(\sum_{b \in \text{VALUES}} P[X=b] \frac{1}{P[X=b]} \right) = \log_2 \left(\sum_{b \in \text{VALUES}} 1 \right) = \log_2 k.$$

with equality only if all the $P[X=b]$ are equal. QED

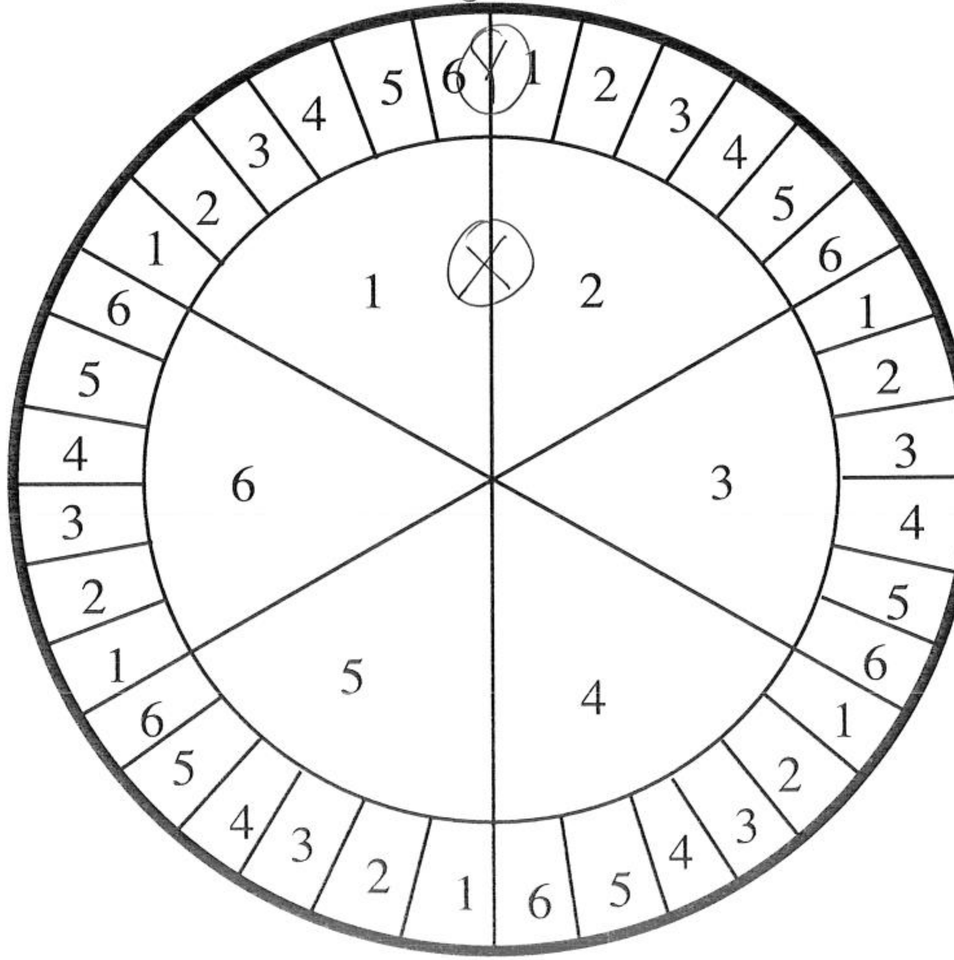


$H(X,Y) \cong 5.17$

$H(X+Y) \cong 3.35$

$H(X,Y \mid X+Y) \overset{?}{=} \cancel{5.17 - 3.35}$ $\cancel{3.35 - 5.17}$

$= 5.17 - 3.35 \simeq 1.82$



$= H(S) + H(R|S)$

$H(R,S) = H(R) + H(S|R)$

$H((X,Y), X+Y) = \cancel{H(\underbrace{X,Y}_{\simeq 5.17}) + \boxed{H(X,Y \mid X+Y)}}$

$\cancel{\underbrace{H(X+Y)}_{\simeq 3.35} + \underbrace{H(X+Y \mid X,Y)}_{=0}}$

$H($

$\underbrace{H(X,Y)}_{\simeq 5.17} + \underbrace{H(X+Y \mid X,Y)}_{=0}$

$= \underbrace{H(X+Y)}_{\simeq 3.35} + \boxed{H(X,Y \mid X+Y)}$

Theorem 3 *For any two random variables X and Y we have*

$$H(X, Y) \leq H(X) + H(Y)$$

with equality holding if and only if X and Y are independent

$H(Y|X) \leq H(Y)$ with equality iff X&Y independ.

Proof. Combining the equality given by Theorem 1 with the inequality of Theorem 2 we get

$$H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y),$$

as desired. Since we have used Theorem 2 we see that equality can only hold true if X and Y are independent. **QED**

Since for a given a, the conditional probabilities $P[Y = b|X = a]$ add up to 1, we can use the convex function inequality

$$\sum_b m_b \log_2 x_b \leq \log_2 \left(\sum_b m_b x_b \right)$$

For appropriate choices of m_b and x_b we have:

$$\begin{aligned} \sum_b P[Y = b|X = a] \log_2 \frac{1}{P[X = a|Y = b]} &\leq \log_2 \left(\sum_b P[Y = b|X = a] \frac{1}{P[X = a|Y = b]} \right) \\ &= \log_2 \left(\sum_b \frac{\cancel{P[X = a, Y = b]}}{P[X = a]} \times \frac{P[Y = b]}{\cancel{P[X = a, Y = b]}} \right) \\ &= \log_2 \left(\sum_b \frac{P[Y = b]}{P[X = a]} \right) \\ &= \log_2 \frac{1}{P[X = a]} \end{aligned}$$

Therefore

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

E	S	P	Y	L	E	T	G	P	D	L	C	P	C	P	D	E	W	P	D	D
F	T																			
G	U	R																		
H	V																			
I	W	T																		
J	X																			
K	Y																			
L	Z																			
M	A	X	G																	
N	B																			
O	C	Z																		
P	D																			
Q	E																			
R	F																			
S	G																			
T	H	E	N	A	T	I	V	E	S	A	R	E	R	E	S	T	L	E	S	S
U	I																			
V	J																			
W	K																			
X	L																			
Y	M																			
Z	N																			
A	O																			
B	P																			
C	Q																			
D	R	O	X																	

Monoalphabetic Substitution

Assume that we have intercepted N letters of a ciphertext message that was encoded using a Monoalphabetic substitution and that the entropy of english is 2 bits.

Length of text	5	10	15	20	30	40	50
# of distinct letters	4	8	11	12	14	16	18

$N \approx$

For instance, a typical english sample of 30 letters contains about 14 di erent letters. Thus the key for a Monoalphabetic substitution only permutes 14 letters. Therefore the number of keys is

$$26 \times 25 \times \cdots \times 13$$

# of Keys for mono alphabetc substitution given that I'm only using 14 distinct letters

and not 26!.

Assuming that each key is equally likely, we have

$$\boxed{H(K) = \log_2(26 \times 25 \times \cdots \times 13) \approx 59.54}$$

Assuming that each of the 26^N ciphertexts is equally likely, we have

$$\boxed{H(C) = \log_2 26^N = N \log_2 26 \approx 4.7N}$$

Therefore,

$$\boxed{59.54 = 4.7N - 2N \Rightarrow N \approx 22.05}$$

Hill 2×2 mod 29

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

if $ad \neq 0$, then
$ad - bc \neq 0$ so $ad \neq bc$

Roughly there are 29^4 possible keys (but there are less because not all 29^4 keys have $ad-bc\neq 0$)

$H(K) = \log_2 29^4$

$H(C) = \log_2(29^N) = N\cdot\log_2 29$ with N letters of cyphertext

$H(M) = 1.2N$ (maybe $2N$)

$H(C) = H(K) + H(M)$

$N\cdot\log_2 29 = 4\cdot\log_2 29 + 1.2N$

$N\cdot 4.9 = 4\cdot(4.9) + 1.2N$

$F=1.2 \quad N = \dfrac{4\cdot(4.9)}{3.7} \simeq 5.3$ (seems small)

$F=3.2 \quad N = \dfrac{4\cdot(4.9)}{1.7} \cong 11.5$