Elements of Information Theory

Definition 1. The **entropy** of an event $A$ is a measure of the uncertainty we feel about the occurrence of $A$.

The entropy of a random variable $X$ is given by

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

Definition 2. The **entropy of two random variables** $X$ and $Y$ is

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

Definition 3. The **conditional entropy of the random variable $X$ given an event $B$** is

$$H(X|B) = \sum_a P(X = a|B) \cdot \log_2 \left( \frac{1}{P(X = a|B)} \right)$$

Definition 4. The **conditional entropy of the random variable $X$ given a random variable $Y$** is

$$H(X|Y) = \sum_b P(Y = b) \cdot H(X|Y = b)$$
Example. Suppose that random variables $X, Y, Z$ are obtained by spinning the wheel below, with $X$ given by the innermost circle, $Y$ given by the intermediate circle, and $Z$ given by the outermost circle.

(a) Compute $H(X)$

(b) How many bits (of information) are required to store the results of 100,000 spins of $Z$?

(c) Calculate the uncertainty of $Z$ given that $X = 0$. 
(d) Calculate $H(Z|Y)$

Let $X, Y$ be random variables. Then

$$H(Y|X) = H(Y) \iff X \text{ and } Y \text{ are independent}.$$
(e) Calculate $H(X|Y, Z)$

Let $X, Y_1, Y_2, \ldots, Y_k$ be random variables then

$$H(X|Y_1, Y_2, \ldots, Y_k) = 0 \iff X = f(Y_1, \ldots, Y_k).$$
Suppose we learn the value of $X$ and then we learn the value of $Y$. Then

$$H(X, Y) = H(X) + H(Y|X).$$
**Important inequalities**

1. 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
2. For any two random variables $X$ and $Y$ we always have

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

and equality holds if and only if $X$ and $Y$ are independent
3. For any two random variables $X$ and $Y$ we always have

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

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