The Entropy of An Event

Definition: The *entropy of an event* A is:

1. the measure of uncertainty we *feel* about the occurrence of A.
2. the amount of *information*, measured in bits, contained by A.

Events that occur with equal probability have the same amount of uncertainty and contain the same amount of information

$\Downarrow$

The entropy of an event should be a function of the probability of that event occurring

The entropy of event $A = h(P(A))$

What properties should the entropy function, h, have to numerically express the measure of our uncertainty about the occurrence of an event in a manner which is compatible with our intuitive notion of uncertainty?

Basic Requirements

1. The more probable the event the smaller the uncertainty

   $h(x)$ should be a decreasing function

2. The uncertainty about the simultaneous occurrence of two independent events is equal to the sum of the individual uncertainties

   $$h(xy) = h(x) + h(y)$$

3. Small changes in the probability should correspond to small changes in the uncertainty

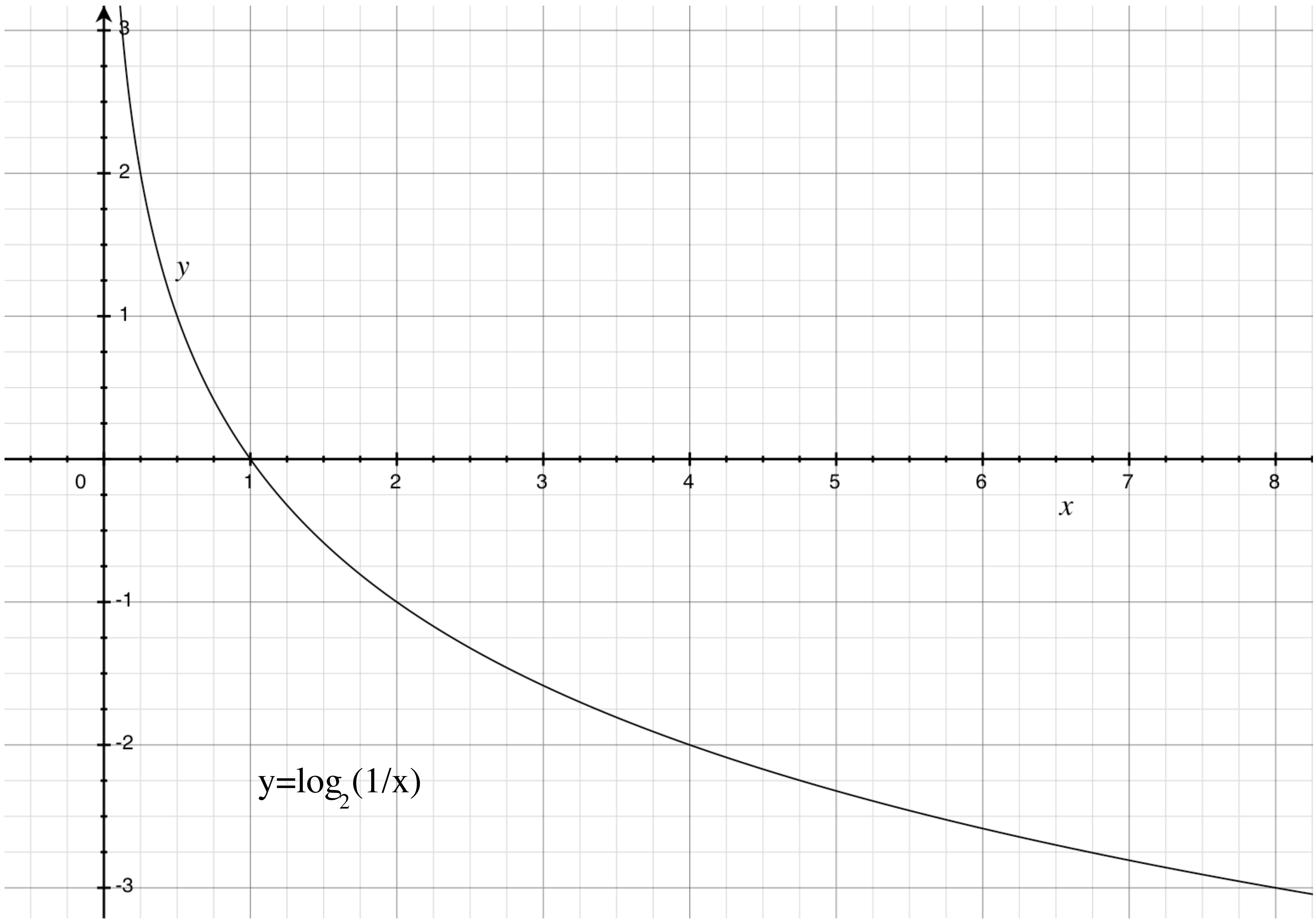
   $h(x)$ should be a continuous function

4. Recording the outcome of a 50/50 situation requires one binary register.

   $$h(1/2) = 1 \text{ (bit)}$$

Therefore

$$h(x) = \log_2 1/x$$



y
x
0
1
2
3
4
5
6
7
8
3
2
1
-1
-2
-3
y=log2(1/x)

Some identities with log

$$\log_b(1) = 0$$

$$\log_b(0) = -\infty \quad \text{(or undefined)}$$

$$\log_b(b) = 1$$

$$\log_b(b^a) = a$$

$$b^{\log_b(a)} = a$$

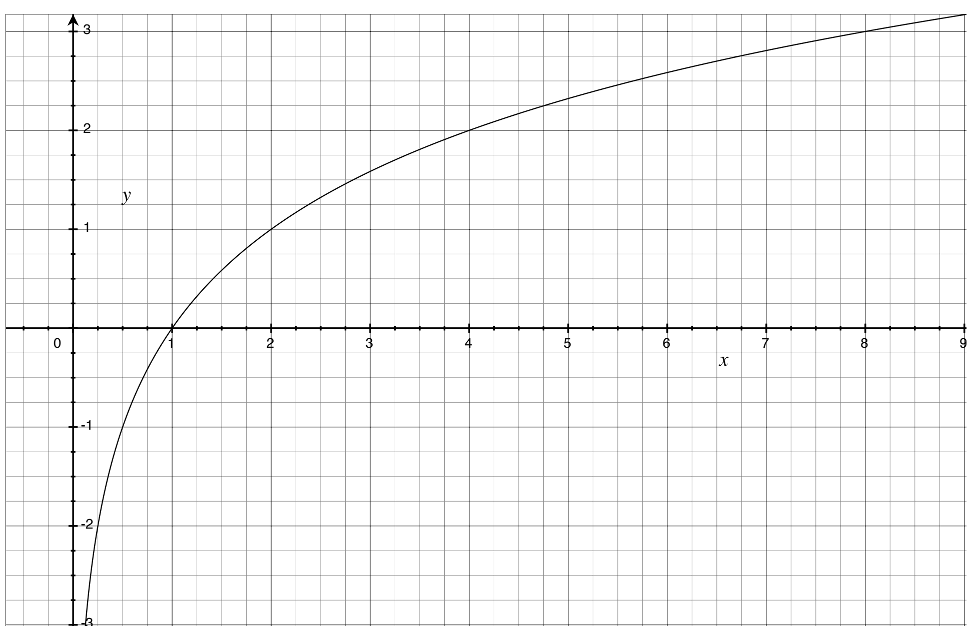
$$\log_2(a) = \frac{\log_b(a)}{\log_b(2)}$$

$$\log(1/a) = -\log(a)$$

$$\log(ab) = \log a + \log b$$

$$\log(a/b) = \log a - \log b$$

$$\log(a^b) = b \log a$$



Graph of $y = \log_2(x)$

Information Theory Definitions

Definition: The Entropy of a random variable X

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

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

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

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

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

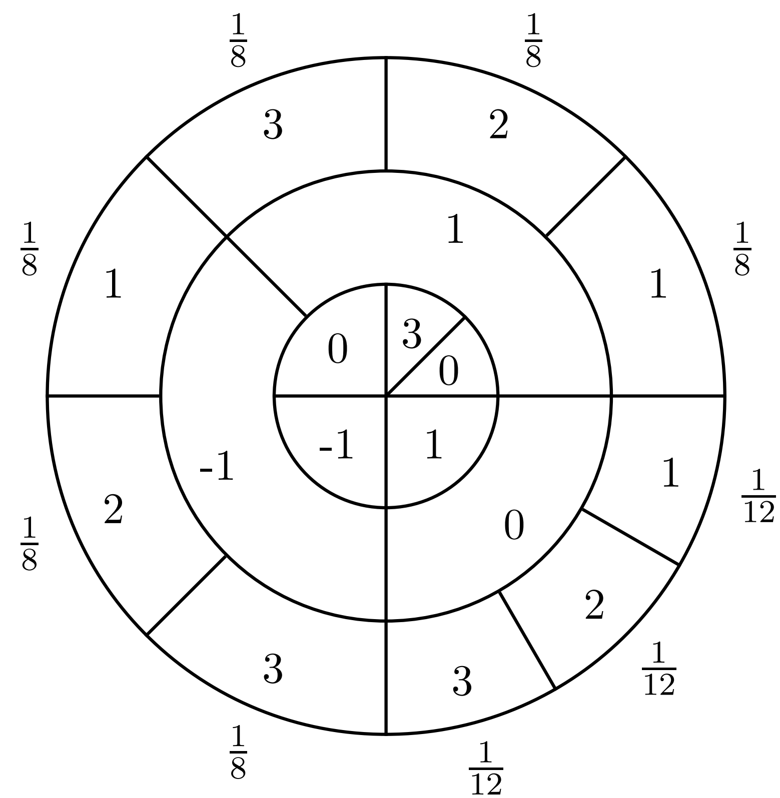
(a) Calculate $H[X]$.

(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 $H[Z|Y]$.



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

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

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

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

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

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

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

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

For appropriate choices of m_b and x_b we have:

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

Therefore

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

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

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**

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

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

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