15 Basic Properties of the Entropy Function

Let X be a discrete random variable taking values over a finite set $A = \{a_1, \cdots, a_n\}$.

Definition 15.1 The Entropy of X denoted by H(X) is defined as

$$H(X) = \sum_x P(X = x) \log \frac{1}{P(X = x)}$$

Here *log* denotes logarithm to base 2 and the sum is assumed to be only over x so that $P(X = x) \neq 0$. Alternatively, we could define $p \log \frac{1}{p} = 0$ for $p = 0$.

Definition 15.2 Given two random variables X and Y, their joint entropy denoted by $H(X, Y)$ is defined as

$$H(X, Y) = \sum_{x,y} P(X = x, Y = y) \log \frac{1}{P(X = x, Y = y)}$$

Let X_y denote the random variable X conditioned on the event $Y = y$. In other words, $P(X_y = x) = P(X = x, Y = y)$. We now define conditional entropy.

Definition 15.3 The entropy of X conditioned on Y, denoted by $H(X|Y)$ is defined as

$$H(X|Y) = E_y(H(X_y))$$

The entropy of X conditioned on Y may be considered to be the entropy that remains in X once the value of Y is known. This is made precise by the following lemma.

Lemma 15.4 $H(X|Y) = H(X, Y) - H(Y)$

Proof.

$$\begin{aligned}
H(X|Y) &= \sum_y P(Y=y) \sum_x P(X_y = x) \log \frac{1}{P(X_y = x)} \\
&= \sum_y P(Y=y) \sum_x \frac{P(X=x, Y=y)}{P(Y=y)} \log \frac{P(Y=y)}{P(X=x, Y=y)} \\
&= \sum_{x,y} P(X=x, Y=y) \log \frac{1}{P(X=x, Y=y)} - \sum_{x,y} P(X=x, Y=y) \log \frac{1}{P(Y=y)} \\
&= \sum_{x,y} P(X=x, Y=y) \log \frac{1}{P(X=x, Y=y)} - \sum_{y} P(Y=y) \log \frac{1}{P(Y=y)} \\
&= H(X,Y) - H(Y)
\end{aligned}$$

□

Corollary 15.5

$$\begin{aligned}
H(X,Y) &= H(X|Y) + H(Y) \\
&= H(Y|X) + H(X)
\end{aligned}$$

Corollary 15.6 $H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$

Proof.

$$\begin{aligned}
H(X,Y|Z) &= E_z(H(X,Y|Z=z)) \\
&= E_z(H(X|Z=z) + H(Y|X,Z=z)) \\
&= E_z(H(X|Z=z)) + E_z(H(Y|X,Z=z)) \\
&= H(X|Z) + H(Y|X,Z)
\end{aligned}$$

□

Corollary 15.5 can be generalized to n variables.

Theorem 15.7 The Chain Rule for Entropy:

$$H(X_1, \cdots, X_n) = H(X_1) + H(X_2|X_1) + \cdots H(X_n|X_1, \cdots, X_{n-1})$$

Proof. Proof by induction. The case $n = 2$ is Corollary 15.5. Assume the statement holds for $n - 1$. To prove the rule for n, consider X_1, X_2 to be a joint random variable and apply the rule for $n - 1$.

$$\begin{aligned}
H(X_1, \cdots, X_n) &= H(X_1, X_2) + H(X_3|X_1, X_2) + \cdots + H(X_n|X_1, X_2, \cdots X_{n-1}) \\
&= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \cdots + H(X_n|X_1, \cdots X_{n-1})
\end{aligned}$$

In the last step, we use corollary 15.5. □

Definition 15.8 The mutual information of X and Y, denoted by $I(X, Y)$ is defined as

$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

Clearly

$$\begin{aligned}
I(X, Y) &= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= \sum_{x,y} P(X = x, Y = y) \log \frac{P(X = x, Y = y)}{P(X = x)P(Y = y)}
\end{aligned}$$

To show that the mutual information is always non-negative, we introduce the notion of relative entropy.

Definition 15.9 The relative entropy of distributions p and q over the same sample space is defined as

$$D(p||q) = \sum_x p_x \log \frac{p_x}{q_x}$$

Theorem 15.10 *$D(p||q) \geq 0$ with equality iff $p_x = q_x$ $\forall x$.*

Proof.

$$\begin{aligned}
-D(p||q) &= \sum_x p_x \log \frac{q_x}{p_x} \\
&\leq \log \left(\sum_x p_x \frac{q_x}{p_x} \right) \qquad (Jensen's\ inequality) \\
&= \log 1 \\
&= 0
\end{aligned}$$

Hence $D(p||q) \geq 0$. Equality holds iff

$$\frac{q_x}{p_x} = \frac{q_y}{p_y} \quad \forall x \neq y$$

But since p and q are both probability distributions, this is possible iff $q_x = p_x \ \forall x$. $\square$

Corollary 15.11 *$I(X,Y) \geq 0$ with equality iff X and Y are independent.*

Corollary 15.12 *$H(X|Y) \leq H(X)$ with equality iff X and Y are independent.*

Corollary 15.13 *If $n = |range(x)|$, $H(X) \leq \log n$ with equality iff X is uniformly distributed.*

Proof. If X is uniformly distributed,

$$H(X) = \sum_{x=1}^{n} \frac{\log n}{n} = \log n$$

Else, denoting the uniform distribution by q and using the fact that $D(p||q) > 0$ if $p \neq q$, we get

$$\begin{aligned}
\sum_i p_i \log\left(\frac{p_i}{q_i}\right) &> 0 \\
\Rightarrow \sum_i p_i \log(np_i) &> 0 \\
\Rightarrow \sum_i p_i \log(n) - \sum_i p_i \log\frac{1}{p_i} &> 0 \\
\Rightarrow \log(n) - \sum_i p_i \log\frac{1}{p_i} &> 0 \\
\Rightarrow H(X) < \log n &
\end{aligned}$$

$\square$

Lemma 15.14 Shearer's Lemma: *Let $X = (X_1, \cdots, X_n)$ be a vector of random variables. Let A be a hypergraph of minimum degree k. For each hyperedge $e \in H$, denote by X_e the vector $\{X_i | i \in e\}$. Then*

$$H(X) \leq \frac{1}{k} \sum_{e \in A} H(X_e)$$

Proof. By the chain rule for entropy

$$H(X) = \sum_{j=1}^{n} H(X_j | X_l,\ l < j)$$

For each edge $e \in A$,

$$\begin{aligned} H(X_e) &= \sum_{j \in e} H(X_j | X_l,\ l < j,\ l \in e) \\ &\geq \sum_{j \in e} H(X_j | X_l,\ l < j) \end{aligned}$$

Adding the equations for each edge, and using the fact that every vertex is contained in at least k edges,

$$\begin{aligned} \sum_{e \in A} H(X_e) &\geq k \sum_{j=1}^{n} H(X_j | X_l,\ l < j) \\ &= kH(X) \end{aligned}$$

□

The above proof of Shearer's lemma is due to Jaikumar Radhakrishnan. We conclude with a simple geometric application.

Lemma 15.15 *If S is a set of n points in R^3 which has n_1 distinct projections on the xy plane, n_2 projections on the xz plane and yz projections on the yz plane. Then $n_1 n_2 n_3 \geq n^2$.*

Proof. Let $P = (X, Y, Z)$ be a random point from S. By Lemma 15.13, $H(P) = \log n$. Define the projection vectors,

$$P_1 = (X, Y)$$
$$P_2 = (X, Z)$$
$$P_3 = (Y, Z)$$

Apply Shearer's lemma to the the hypergraph with vertices $[1, 2, 3]$ and edges $\{(1, 2), (2, 3), (1, 3)\}$. Clearly $k = 2$.

$$\begin{aligned} 2H(P) &\leq H(P_1) + H(P_2) + H(P_3) \\ &\leq \log(n_1) + \log(n_2) + \log(n_3) \\ \Rightarrow 2\log n &\leq \log(n_1 n_2 n_3) \\ \Rightarrow n^2 &\leq n_1 n_2 n_3 \end{aligned}$$

$\square$