EE/Ma 126a Information Theory - Homework Set #1

Ling Li, ling@cs.caltech.edu

October 3, 2000

1.1 Coin flips

X is positive integer. For $X = n$, which means the coin is flipped n times and only the last flip is head, the probability is

$$p(X = n) = 2^{-n},$$

since the coin is fair. Then we have

$$\boxed{H(X) = -\sum_{n=1}^{\infty} p(n) \log p(n) = \sum_{n=1}^{\infty} 2^{-n} \cdot n = 2}$$

In order to find an "efficient" sequence of yes-no questions, we should begin with questions about numbers with the most probability. That is, we should first ask "Is X contained in $\{1\}$?", then "Is X contained in $\{2\}$?", and so on. For $X = n$, n questions would be asked before we get X. So, the expected number of questions required to determine X is

$$\boxed{\sum_{n=1}^{\infty} n \cdot p(n) = \sum_{n=1}^{\infty} n \cdot 2^{-n} = 2}$$

which *equals* $H(X)$.

1.2 Entropy of functions

Note that entropy is a functional of the distribution not depending on the actual values the random variable takes.[1] So for (a) $Y = 2^X$, the distribution doesn't change since $2^{x_1} = 2^{x_2} \Leftrightarrow x_1 = x_2$. Thus $\boxed{H(Y) = H(X).}$

As for (b) $Y = \cos X$, different values of X may result in sames value of Y, for $\cos(\cdot)$ is a periodical function. Then the distribution of Y may be 'condensed', which means $\boxed{H(Y) \le H(X).}$ From Problem 1.4, we also know that $H(Y) \le H(X)$.

1.3 Minimum entropy

Entropy is non-negative. So if some $\mathbf{p}$ makes $H(\mathbf{p}) = 0$, that is the minimum entropy. Since

$$H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log p_i$$

and $p_i \log p_i \leq 0$, we should find some $\mathbf{p}$ so that $p_i \log p_i = 0$ for all $i = 1..n$. This means $p_i = 0$ or $p_i = 1$ for all $i = 1..n$. Considering with $\sum_{i=0}^{n} p_i = 1$, p_i should be all zero except one is 1. So, the $\boxed{\text{minimum value of } H(\mathbf{p}) \text{ is } 0,}$ and $\mathbf{p}$ should be one of

$$\boxed{(1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots 0, 1)}$$

1.4 Entropy of functions of a random variable

For all $x \in \mathcal{X}$, $X = x \Rightarrow g(X) = g(x)$. So $p(g(X)|X) = 1$,

$$H(g(X)|X) = E \log \frac{1}{p(g(X)|X)} = 0.$$

From the chain rule ([1, theorem 2.2.1]), we know that

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

Since entropy is non-negative, we also get[1]

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

Thus $H(g(X)) \leq H(X)$. $\square$

1.5 Zero conditional entropy

We know that $\log p(y|x) \leq 0$ and $p(x, y) \geq 0$, and

$$H(Y|X) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) = 0.$$

So $p(x, y) > 0 \Rightarrow \log p(y|x) = 0$, i.e., $p(y|x) = 1$. Thus for each x with $p(x) > 0$, if some y makes $p(x, y) > 0$, that y is unique since $p(y|x) = 1$. In other word, for all x with $p(x) > 0$, there is only one possible value of y with $p(x, y) > 0$. That means Y is a function of X. $\square$

References

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[1] g may not be a reversable function, which means for some $X = x$, $0 < p(x|g(x)) < 1$. Thus $H(X|g(X)) > 0$.