CS909 Topics In Algorithms	October 28, 2003
Lecture 25	
Lecturer: Amit Kumar	*Scribe: Neha Jain*

1 Introduction

In the next few lectures, we shall give an introduction to information theory. Information theory gives us a measure of the information contained in a random variable. Consider an experiment such that the outcome of the experiment is a random variable. A particuler outcome of the experiment has some amount of information associated with its occurence. The potential information which can be conveyed by the experiment can be defined as the entropy of the experiment.

First we shall explain the entropy function associated with an event of the experiment. Then we shall derieve an expression for the entropy function of an experiment. We shall conclude with the proof of two lemmas which shall be used in subsequent lectures.

2 Information of an Event

Consider an event X. Denote the probability of occurance of X as p_x. Consider the case when $p_x < 1$. This means that before X occurs, we are not sure wether X would occur or not. When finally the experiment happens, we get to know whether X occured or not. Its occurance is an information for us, and the information content is higher if the probability of occurance of X is smaller. In fact if $p_x = 1$, then the occurance of X does not have any information.

Consider the following two experiments

Tossing of a unbiased coin : There are two possible outcomes : {Heads, Tails}. Each event is equally likely. Thus before tossing the coin we have no idea what the result could be. When we finally toss the coin, whatever the outcome is, it is an information for us. And both the outcomes convey equal amount of information.

Tossing a biased coin : proabability of a Heads is 0.9, then the occurance of the Heads conveys less information than the occurance of Tails, since we were already expecting a Heads with greater probability.

Thus information of an event is a function of the probability of occurance of the event. For the rest of our discussion, we shall denote the information of X as $I(p_x)$.

3 Entropy of an Experiment

The entropy of an experiment tries to capture the expected amount of information conveyed by the experiment.

Consider an experiment with two possible outcomes x, x'. The probability of occurance of these events is p_x and $p_{x'}$, respectively. We want to define a function which gives a measure of the expected information that the experiment can convey. Note that if x occurs, then the measure of the information we get is $I(p_x)$. But the probability that we get this much information is equal to the probability of occurance of x which is p_x. Thus we can define the *entropy* of this experiment as

$$H(x, x') = p_x I(p_x) + p_{x'} I(p_{x'})$$

3.1 General Exprression for Entropy

Consider an experiment with sample space : $\{x_1, x_2, \ldots, x_n\}$.

- Probability of occurance of x_i is p_i.
- Since these are the only possible events, $p_1 + p_2 + \cdots + p_n = 1$.

We want to define an entropy function for this experiment, $H(p_1, p_2, \ldots, p_n)$ such that it satisfes the following :

1. $H(p_1, p_2, \ldots, p_n)$ captures the expected information of the experiment.
$$H(p_1, p_2, \ldots, p_n) = \sum p_i I(p_i)$$
2. $H(p_1, p_2, \ldots, p_n)$ is symmeteric.
If $p'_1, p'_2, \ldots, p'_n$ is a reordering of $p_1, p_2, \ldots, p_n$, then
$$H(p_1, p_2, \ldots, p_n) = H(p'_1, p'_2, \ldots, p'_n)$$
3. $H(p, 1-p)$ is a continuous function of p.
4. Conservation of Entropy
Consider the case when $n \geq 3$. Let $\bar{x_1}$ denote the event that x_1 does not occur.
$$\mathbf{Pr}(\bar{x_1}) = \bar{p_1} = 1 - p_1$$
For $i \neq 1$, we can define
$$\mathbf{Pr}(x_i|\bar{x_1}) = \frac{p_i}{\bar{p_1}}$$
$$H\left(\frac{p_2}{\bar{p_1}}, \frac{p_3}{\bar{p_1}} \ldots, \frac{p_n}{\bar{p_1}}\right) = \sum_{i=2}^{n} \frac{p_i}{\bar{p_1}} I\left(\frac{p_i}{\bar{p_1}}\right)$$
$$\Rightarrow H(p_1, p_2 \ldots, p_n) = H(p_1, \bar{p_1}) + (1-p_1) H(\frac{p_2}{\bar{p_1}}, \frac{p_3}{\bar{p_1}} \ldots, \frac{p_n}{\bar{p_1}})$$

The only solution that satisfies the four conditions above is

$$H(p_1, p_2 \ldots, p_n) = c \sum_i p_i \log(\frac{1}{p_i})$$

where c is a constant. We will fix c to be 1, and the logarithm will be with respect to base 2.

3.2 Verification of the Conditions above

3.2.1 Condition 1

$$H(p_1, p_2 \ldots, p_n) = \sum_i p_i \log(\frac{1}{p_i}) = \sum_i p_i I(p_i)$$

This gives us $I(p_i) = \log(\frac{1}{p_i})$. Recall from our discussion in section 2 that we want that

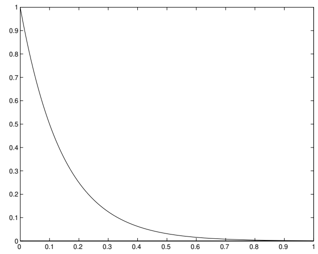

Figure 1: The plot of $I(p)$ vs p

$I(1)$ should be 0 and $I(p_i)$ should increase for decreasing values of p_i. The function $\log(\frac{1}{p_i})$ satisfies this property as can be seen from the graph of the function in 1

3.2.2 Condition 2

$$H(p, 1-p) = -p\log(p) - (1-p)\log(1-p)$$

We define $\lim_{p \to 0} p\log(p) = 0$ The contitnuity of $H(p, 1-p)$ can be verified from the graph of the function given in 2

3.2.3 Condition 4

We want to show that

$$H(p_1, p_2 \ldots, p_n) = H(p_1, p'_1) + (1-p_1)H(\frac{p_2}{p'_1}, \frac{p_3}{p'_1} \ldots, \frac{p_n}{p'_1})$$

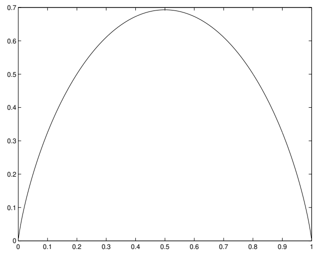

Figure 2: The plot of $H(p, 1-p)$ vs p

Evaluating the left hand side of this equation

$$= -p_1 \log(p_1) + (1-p_1)\log(1-p_1) - (1-p_1)\sum_{i=2}^{n} \frac{p_i}{1-p_1} \log(\frac{p_i}{1-p_1})$$

$$= -p_1 \log(p_1) + (1-p_1)\log(1-p_1) - \sum_{i=2}^{n} p_i \log(p_i) + \sum_{i=2}^{n} p_i \log(1-p_1)$$

$$= -p_1 \log(p_1) - (1-p_1)\log(1-p_1) - \sum_{i=2}^{n} p_i \log(p_i) + \log(1-p_1)(1-p_1)$$

$$= -p_1 \log(p_1) - \sum_{i=2}^{n} p_i \log(p_i) = -\sum_{i=1}^{n} p_i \log(p_i) = H(p_1, p_2, \ldots, p_n)$$

Hence Proved

4 Lemmas

Lemma 25.1 Given $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$ such that

$$\sum_i p_i = 1 \qquad p_i \geq 0$$

$$\sum_i q_i = 1 \qquad q_i \geq 0$$

Then

$$-\sum_i q_i \log(q_i) \leq -\sum_i q_i \log(p_i)$$

with equality iff $q_i = p_i, \forall_i$

Proof: Using Jensen's inequality[1]

$$\prod_i \frac{p_i}{q_i}^{q_i} \leq \sum_i \frac{p_i}{q_i} q_i \leq 1$$

[1] Given $a_1, a_2, \ldots, a_n$ and $x_1, x_2, \ldots, x_n$ such that $\sum_i x_i = 1$

$$\prod_i a_i^{x_i} \leq \sum_i a_i x_i$$

Taking log on both sides, we get

$$\sum_i q_i \log(p_i) - q_i \log(q_i) \leq 0$$

$$\Rightarrow \sum_i q_i \log(p_i) \leq \sum_i q_i \log(q_i)$$

Taking negative of both sides and reversing the signs of the inequality, we get

$$-\sum_i q_i \log(q_i) \leq -\sum_i q_i \log(p_i)$$

□

Lemma 25.2 Given a vector $p = [p_1, p_2, \ldots, p_n]$ such that

$$\sum_j p_j = 1 \qquad\qquad p_j \geq 0$$

and a non negative matrix $A = [a_{ij}]$ such that $a_{ij} \geq 0$ and

$$\sum_i a_{ij} = 1 \qquad\qquad \sum_j a_{ij} = 1$$

Define

$$p'_i = \sum_j a_{ij} p_j$$

then

$$-\sum_i p'_i \log(p'_i) \geq -\sum_i p_i \log(p_i)$$

with equality iff p'_i is a reordering of p_i

Proof: We have $p'_i = \sum_j a_{ij} p_j$. This gives us

$$\begin{aligned} -\sum_i p'_i \log(p'_i) &= -\sum_i \sum_j a_{ij} p_j \log(p'_i) \\ &= -\sum_j p_j \sum_i a_{ij} \log(p'_i) \\ &= -\sum_j p_j \log(\prod_i p_i'^{a_{ij}}) \end{aligned}$$

By Jensen's inequality, we have

$$\begin{aligned} \prod_i p_i'^{a_{ij}} &\leq \sum_i p_i a_{ij} \\ \Rightarrow -\sum_j p_j \log(\prod_i p_i'^{a_{ij}}) &\geq -\sum_j p_j \log(\sum_i p_i a_{ij}) \\ \Rightarrow -\sum_i p_i \log(p'_i) &\geq -\sum_j p_j \log(\sum_i p_i a_{ij}) \end{aligned}$$

Take $r_j = \sum_i p_i a_{ij}$. We know that $\sum_j r_j = \sum_j \sum_i a_{ij} p_i = 1$

Using lemma 25.1, we get

$$-\sum_j p_j \log(r_j) \geq -\sum_j p_j \log(p_j)$$

Thus we can conclude

$$-\sum_j p'_j \log(p'_j) \geq -\sum_j p_j \log(r_j) \geq -\sum_j p_j \log(p_j)$$

$\square$