Shannon Entropy

Satyadev Nandakumar

December 29, 2013

In this section, we will give the definitions of Shannon entropy, motivating the definition from multiple perspectives. We will use a finite alphabet X containing t symbols.

Definition 0.1. Let P be a probability distribution on X. Then the entropy of P is defined as

$$H(P) = -\sum_{a \in X} P(a) \log P(a).$$

1 Combinatorial Motivation

The *type* or *empirical distribution* of a string $x \in X^n$ is the distribution on the symbols in X determined by

$$P_x(a) = \frac{\text{number of times } a \text{ occurs in } x}{n} \qquad (a \in X).$$

For example, if X is the binary alphabet, the type of 0111 is

$$P_{0111}(0) = \frac{1}{4},\ P_{0111}(0) = \frac{3}{4}.$$

A distribution P is called an n-type if there is a string with that distribution. The set of strings of a particular type P_x is denoted T^n_P. The superscript here denotes that this set consists of n-long strings.

For example, if P is the distribution $P(0) = \frac{1}{4}, P(1) = \frac{3}{4}$, then $T^4 = \{0111, 1011, 1101, 1110\}$. An example of a distribution on the binary alphabet that cannot be a 4-type is $P(0) = \frac{1}{7}, P(1) = \frac{6}{7}$, since no 4-long string can have such an empirical distribution.

Lemma 1.1. *The number of possible n-types defined on X is*

$$\binom{n + |X| - 1}{|X| - 1}.$$

Proof. An n-type can be defined by considering a sequence of n ones, and inserting $|X|-1$ partitions between them. Any such partition defines a unique type. The number of such partitions of n is

$$\binom{n + |X| - 1}{|X| - 1}.$$

□

We now try to show that there is a natural question which we can ask, whose answer will lead us to Shannon entropy. The question is of fairly tight upper and lower bounds for the number of elements of any type.

Lemma 1.2. *For any n-type P,*

$$\frac{2^{nH(P)}}{\binom{n+|X|-1}{|X|-1}} \quad \leq |T_P^n| \leq \quad 2^{nH(P)}.$$

Proof. Suppose P is defined by

$$P_x(a_i) = \frac{k_i}{n}, \qquad 1 \leq i \leq |X|,$$

where $X = \{a_1, a_2, \ldots, a_t\}$ and $\sum_{i=1}^t k_i = n$.

We know that

$$|T_P^n| = \frac{n!}{k_1!k_2!\ldots k_t!}.$$

This multinomial coefficient suggests that we can get an estimate at the above term by looking at multinomial expansions. We have

$$n^n = (k_1 + k_2 + \cdots + k_t)^n = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_{t-1}=1}^{n} \frac{n!}{j_1!j_2!\ldots j_{t-1}!j_t!} k_1^{j_1} k_2^{j_2} \ldots k_t^{j_t},$$

where $\sum_{i=1} tj_i = n$. Let us denote the summands as

$$S[j_1, j_2, \ldots, j_t].$$

The number of summands is $\binom{n+|X|-1}{|X|-1}$.

The largest of the summands is the term

$$\frac{n!}{k_1!k_2!\ldots k_{t-1}!k_t!} k_1^{k_1} k_2^{k_2} \ldots k_t^{k_t},$$

by the following consideration. Suppose at some index i, $j_i > k_i$. Necessarily, there must exist another index m where $j_m < k_m$. Without loss of generality, suppose $j_1 > k_1$, and $j_2 < k_2$. Then the ratio

$$\frac{S[j_1, j_2, \ldots, j_t]}{S[j_1 - 1, j_2 + 1, \ldots, j_t]} = \frac{j_1 k_2}{j_2 k_1} < 1,$$

so the maximal summand is $S[k_1, k_2, \ldots, k_t]$.

There is a correspondence between the summands and the n-types. Thus an upper bound on the sum is

$$n^n \leq S[k_1 k_2 \ldots k_t] \binom{n + |X| - 1}{|X| - 1} = \frac{n!}{k_1!k_2!\ldots k_{t-1}!k_t!} k_1^{k_1} k_2^{k_2} \ldots k_t^{k_t} \binom{n + |X| - 1}{|X| - 1} = \frac{n!}{k_1!k_2!\ldots k_{t-1}!k_t!} k_1^{k_1} k_2^{k_2} \ldots k_t^{k_t} \, |T$$

For the lower bound on $|T_P^n|$, we observe from the above inequality that

$$\begin{aligned} |T_P^n| \binom{n + |X| - 1}{|X| - 1} &\geq \frac{n^n}{\Pi_{i=1}^t (k_i)^{k_i}} \\ &= \frac{1}{\Pi_{i=1}^t (\frac{k_i}{n})^{k_i}} \\ &= \Pi_{i=1}^t P_i^{-nP_i} \\ &= \Pi_{i=1}^t 2^{-nP_i \log_2 P_i} \\ &= 2^{\sum_{i=1}^t -nP_i \log_2 P_i} = 2^{nH(P)}. \end{aligned}$$

For the upper bound, since the maximal term is merely one summand, we get that

$$|T_P^n| \leq 2^{nH(P)}.$$

□

Thus, $H(P)$ is a good estimate of $\frac{\log_2 |T_P^n|}{n}$. We can interpret it as the average number of bits used to represent the cardinality of $|T_n^P|$, the averaging being done over the length of the sample, n.