**Lecture 14 ENTROPY**

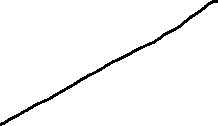
**[Ref #4]**

We now try to *quantify* the amount of uncertainty in a random experiment (or a sequence of identical random experiments) with discrete outcomes.

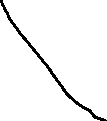
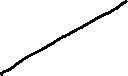
Let S be the sample space of the experiment, and let Ak, k = 1 ... N, be possible disjoint events, such that their union equals S. Recall that an event is defined as a set of outcomes; that is, each Ak is a subset of S.

A1 A2

S



A3 A4



Here we say { A1 ... AN } is a *partition* of S, and we denote p(Ak) by pk. Let U denote this partition.

Clearly k=1..N  pk = 1.

Let H(U) denote the uncertainty associated with partition U of S.

Why do we associate uncertainty with partition U rather than with sample space S? In fact we can say right away that , if N = 1, then H( {S} ) = 0. **Why?**

We must *define* H(U) in a way which is logical, consistent with the axioms of probability – and in agreement with our intuitive understanding of uncertainty.

We postulate that:

1. H(U) is a continuous function of the probabilities pk.

2. If p1 = p2 = ... = pN , then H(U) is an increasing function of N.

3. If partition B is obtained by *refining* partition U, then H(B) > H(U).

The following can be proven:

H(U) as defined below

H(U) = k=1..N  - pk log2(pk) [14.1]

satisfies the above axioms (even when it is multiplied by a constant *c*).

[Strictly speaking, the base of the logarithm in 14.1 need not be 2. Why?]

But 14.1 looks *exactly* like the definition of *information* [contained in an N-ary symbol whose values occur with probabilities pk, k = 1..N]. [Ref. Claude Shannon]

So what is the relationship between ***entropy*** and ***information***?

Hint: A vessel has *capacity* 5 litres. It takes 5 litres water to fill it.

**Example 1**

S = { 0, 1 }

A1 = { 0 }, A2 = { 1 }, partition U = { A1 , A2 }

N = 2

p1 = p2 = 0.5

H(U) = -0.5\*log20.5 - 0.5\*log20.5 = **1 bit**

**Example 2**

S = { 1, 2, 3, 4, 5, 6 }

A1 = { 1 } .... A6 = { 6 }, partition U = { A1, A2, A3, A4, A5,A6 }

N = 6

p1 = ... = p6  = 1/6

H(U) = 6\*[ -1/6\*log2(1/6) ] = log26 **~** **2.585 bits**

**Example 3**



S = { a, b, c .... z } lower case letters, |U| = 26

A1 = { a }, ..... A26 = { z }

Assume that all the letters are equi-probable.

H(U) = 26\*[ -1/26\*log2(1/26) ] = log226 **~ 4.700 bits**

Useful table of ~ log2 values

N Log2(N)

2 1.000

3 1.585

5 2.322

7 2.807

11 3.459

13 3.700

17 4.087

19 4.248

**Example 4 [DIY]**

Similar to the previous one, but it is given that 14 of the 26 letters are TWICE as probable as the remaining 12 letters.

**Example 5**

S = pairs of numbers thrown by a pair of distinct, fair dice

|S| = 36

|U| = 36

Each possible outcome ( j, k ) treated as defining a partition.

H(U) = log236 **~ 2\*2.585 = 5.170 bits**

**Example 6**

S = numbers thrown by the two dice are **ADDED**

|S| = 11

S is partitioned into 11 subsets / events A2 ... A12, so |U| = 11.

Note that these 11 events are NOT equi-probable. Calculate H(U).



|  |  |  |
| --- | --- | --- |
| **i** | **Prob(i)\*36** | **Entropy** |
| 2 | 1 | 0.1436090 |
| 3 | 2 | 0.2316625 |
| 4 | 3 | 0.2987469 |
| 5 | 4 | 0.3522139 |
| 6 | 5 | 0.3955551 |
| 7 | 6 | 0.4308271 |
| 8 | 5 | 0.3955551 |
| 9 | 4 | 0.3522139 |
| 10 | 3 | 0.2987469 |
| 11 | 2 | 0.2316625 |
| 12 | 1 | 0.1436090 |
|  | **Entropy =** | **3.274402** |
|  |  |  |
|  |  |  |
|  | **log(11,2) =** | **3.459432** |

**Is it correct to say that: ENTROPY = UNCERTAINTY = INFORMATION?**

**See comment above.**