Information Theory
University of Cambridge

Jasmeet Bhatia

Contents

Chapter 1

Lectures 1 and 2

1.1 Introduction to Information Theory

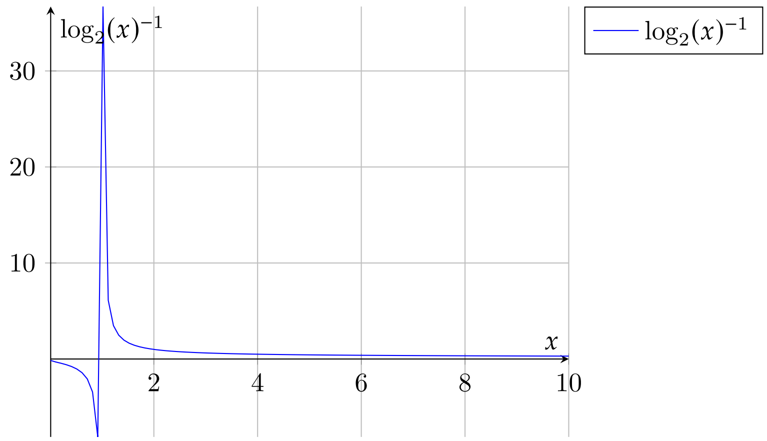
Definition 1.1.1: Shannon's Definition of Information (h)

Let an ensemble, X, have input (x, A_x, P_x) where x is a random variable, A_x is a set of possible outcomes $(a_1, a_2, a_3, ...a_1)$ and where P_x is the set of possibilities $P_x = (P_1, P_2, ...P_i)$ such that P(x=a_i) = P_i, we define Shannon's information content as the following:

$$x = a_i h(x = a_i) = log_2$$

$$\therefore h(x = a_i) = log_2 \frac{1}{P(x = a_i)}$$

1.2 Shannon Plot of Information



Claim 1.2.1 Additive Property

h is additive for independent random variable

Example 1.2.1 (XY)

$x, y = P(x, y) = P(x)P(y)$ and x, y. Then $h(x, y) = log_2 \frac{1}{P(x,y)} = log_2 \frac{1}{P(x)} + log_2 \frac{1}{P(y)}$. Based on this definition and graph above, we see that the Shannon information thoery shows that the h is largest for outcomes that are the most improbable.

1.3 Entropy H(x)

Definition 1.3.1: Entropy H(X)

Entropy of an ensemble is the average Shannon Information Content (h)

$$H(x) = -\sum P(X) log_2 \frac{1}{P(x)}$$

Note:-

Entropy H(x) is in units of bits

Example 1.3.1 (Applying Entropy to Weighing Ball Problem)

Comparing 6 Balls against 6 Balls with 12 Balls Total

$$H(x) = 0.5 log_2 \frac{1}{0.5} = 1 bit$$

Comparing 5 balls against 5 balls with 2 balls on the table

$$H(x) = (1/6) log_2 \frac{1}{12/2} +$$

	Left Ball is Heavier	**Balls is Balanced**	**Right Ball is Heavier**
Probability	1/6	4/6	1/6
-p * log2(p)	-(1/6) * log2(1/6)	-(4/6) * log2(4/6)	-(1/6) * log2(1/6)

Based on this,

$$H(x) = -(1/6) log_2(1/6) - (4/6) log_2(4/6) - (1/6) log_2(1/6) = 1.251 bits$$

Chapter 2

Lecture 3

Claim 2.0.1 Source Coding Theorem
N outcomes from a source X can be compressed into roughly $NH(X)$ bits.

Example 2.0.1 (Bent Coin Lotter)
A coin with p_1 will be tossed $N = 1000$ times. The outcome is x = $x_1, x_2, ...x_n$ e.g x = 00000100100...00010 You can buy 2^n possible tickers for $1 each, before the coin-tossing. If you own ticket x, you win $1,000,000,000.

Question 1: problem

To have a 99% chance of winning, at lowest possible cost, which tickets would you buy? And how many tickets is that? Express your answer in the form $2^{(...)}$

Example 2.0.2 (Sixty-Three)
$x \in \{0, 1, 2, 3, 4, 5, 6..., 63\}$
How can one play the game? One strategy is to split the numbers in half and ask questions that give the most information.
$X\%32 \geqslant 16$?
$X\%16 \geqslant 8$?
$X\%8 \geqslant 4$?

$$h(c_i) = log_2 \frac{1}{P(c_i)}(2) = 1bit$$

Total Shannon Information Content gained = 6 bits. The string $c_1...c_6$ is an encoding of x, defined as c(x) where c(42) - 10100.

Question 2: Generalize

If there are S possible outcomes, how many bits long must each name be, if each outcome has a unique name? Answer: $\lceil log_2(S) \rceil$ An outcome from a set of size S can be communicated in $\lceil log_2(S) \rceil$ bits

2.1 Submarine

Example 2.1.1 (Submarine)

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

$$P(r_1 = n) = \frac{63}{64}$$

= 0.0227 bits

$$P(r_1 = n) = \frac{63}{64}$$

= 0.0227 bits

2.2 Bent Coin Lottery

Example 2.2.1 (Bent Coin Lottery)

A coin with p_1 will be tossed N = 1000 times. The outcome is x = $x_1, x_2, ...x_N$. e.g. x = 000001001000100...00010 You can buy any of the 2^N possible tickets for $1 each, before the coin tossing. If you own ticket x, you win $1,000,000,000.

Question 3: 1

If you are forced to buy one ticket, which would you buy? The all zeroes ticket has the highest chance

Question 4

To have a 99% chance of winning at lowest possible cost, which tickets would you buy? And how many tickets is that? Express your answer in the form $2^{(...)}$

Number of tickets (in order to acquire 99% probability to win) = $1 + 1000 + \binom{1000}{2} + ... + \binom{1000}{101} + \binom{1000}{102} + ... + \binom{1000}{122} + \binom{1000}{123}$

Chapter 3

Lecture 4

3.1 File Compression

Claim 3.1.1 Source Coding Theorem
N outcomes from a source X can be compressed into roughly $NH(X)$ bits.

Proved by counting the typical set When a source X produces N indpendent outcomes
$x = x_1x_2...x_N$ this string is very likely to be one of the $2^{NH(X)}$ typical outcomes all of which have probability $2^{-NH(X)}$