# EES452 2020/2 Part I.3 Dr.Prapun

## 2.4 (Shannon) Entropy for Discrete Random Variables

Entropy is a measure of uncertainty of a random variable [5, p 13].

Entropy quantifies/measures the amount of uncertainty a RV has. (randomness)

It arises as the answer to a number of natural questions. One such question that will be important for us is "What is the average length of the shortest *description* of the random variable?"

expected length of optimal code

**Definition 2.41.** The **entropy** $H(X)$ of a discrete random variable $X$ is defined by

negative sign

Recall: $\log_2 a = \frac{\ln a}{\ln 2} = \frac{\log_{10} a}{\log_{10} 2}$

$$H(X) = -\sum_{x \in S_X} p_X(x) \log_2 p_X(x) = -\mathbb{E}[\log_2 p_X(X)]. = \mathbb{E}[i(X)]$$

$i(x) \equiv -\log_2 p_X(x)$ = the amount of information associated with each realized value of $x$

Usually, I will omit this part. The x value outside the support will give $0 \log_2 0 = 0$ anyway.

Recall $\mathbb{E}[g(x)] = \sum_x p_X(x) g(x)$

- The log is to the base 2 and entropy is expressed in bits (per symbol).
  - The base of the logarithm used in defining $H$ can be chosen to be any convenient real number $b > 1$ but if $b \neq 2$ the unit will not be in bits.
  - If the base of the logarithm is $e$, the entropy is measured in nats.
  - Unless otherwise specified, base 2 is our default base.
- Based on continuity arguments, we shall assume that $0 \ln 0 = 0$.

$0 \log_2 0 = 0$

$\lim_{x \to 0} x \ln x = 0$

Back then, the probability values are $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$

**Example 2.42.** The entropy of the random variable $X$ in Example 2.31 is 1.75 bits (per symbol).

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} = \frac{1}{2}+\frac{1}{2}+\frac{3}{8}+\frac{3}{8}$$

$(-1)\log_2 2 \quad (-2)\log_2 2 \quad (-3)\log_2 2$

$$= 1 + \frac{3}{4} = 1.75 \text{ bits}$$

**Example 2.43.** The entropy of a fair coin toss is 1 bit (per toss).

$$p_X(x) = \begin{cases} 1/2, & x = H, T, \\ 0, & \text{otherwise.} \end{cases}$$

The probabilities involved are $\frac{1}{2}, \frac{1}{2}$

$$H(X) = -2\left(\frac{1}{2}\log_2\frac{1}{2}\right) = 1 \text{ bit}$$

**2.44.** Note that entropy is a functional of the (unordered) probabilities from the pmf of $X$. It does not depend on the actual values taken by the random variable $X$, Therefore, sometimes, we write $H(p_X)$ instead of $H(X)$ to emphasize this fact. Moreover, because we use only the probability values, we can use the row vector representation $\underline{\mathbf{p}}$ of the pmf $p_X$ and simply express the entropy as $H(\underline{\mathbf{p}})$.

In MATLAB, to calculate $H(X)$, we may define a row vector pX from the pmf $p_X$. Then, the value of the entropy is given by

```
HX = -pX*(log2(pX))'.
```

**Example 2.45.** The entropy of a uniform (discrete) random variable $X$ on $\{1, 2, 3, \ldots, n\}$:

$$p_X(x) = \begin{cases} 1/n, & x = 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

$$H(X) = -n\left(\frac{1}{n}\log_2\frac{1}{n}\right) = \log_2 n = \log_2 |S_X|$$

$$\underline{p} = \left[\frac{1}{n}\ \frac{1}{n}\ \cdots\ \frac{1}{n}\right] \quad (n \text{ values})$$

Alternatively,

$$H(X) = -\mathbb{E}\left[\log_2 \underbrace{p_X(X)}_{1/n}\right] = -\mathbb{E}\left[\log_2\frac{1}{n}\right] = -\log_2\frac{1}{n} = \log_2 n$$

**Example 2.46.** The entropy of a Bernoulli random variable $X$:

$$p_X(x) = \begin{cases} p, & x = 1, \\ 1-p, & x = 0, \\ 0, & \text{otherwise.} \end{cases}$$

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p)$$

$$\underline{p} = [p \quad 1-p]$$

Ex. $p = 0.3 \Rightarrow H(X) \approx 0.8813$

$p = 0.5 \Rightarrow H(X) = 1$

Binary RV

$$p_X(x) = \begin{cases} p, & x = b, \\ 1-p, & x = a, \\ 0, & \text{otherwise} \end{cases}$$

same

**Definition 2.47.** **Binary Entropy Function** : We define $h_b(p)$, $h(p)$ or $H(p)$ to be $-p\log_2 p - (1-p)\log_2(1-p)$, whose plot is shown in Figure 3.

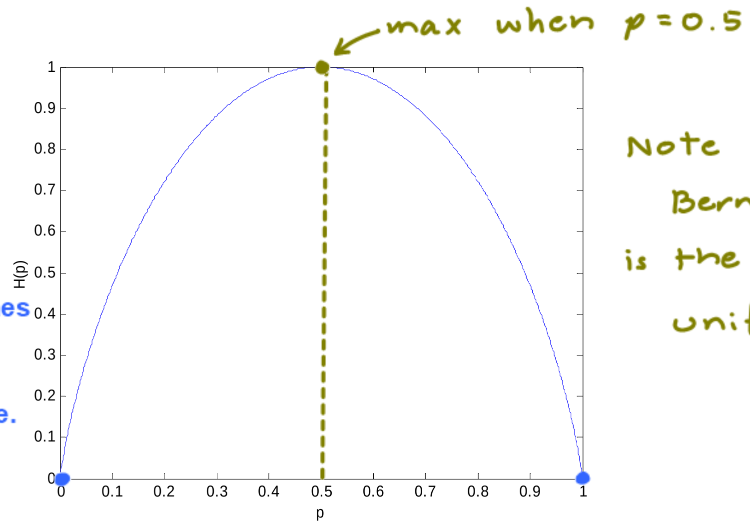


Figure 3: Binary Entropy Function

**2.48.** Two important facts about entropy:

(a) $H(X) \le \log_2 |S_X|$ with equality if and only if $X$ is a uniform random variable.

(b) $H(X) \ge 0$ with equality if and only if $X$ is not random.

In summary,

$$\underbrace{0}_{\text{deterministic}} \le H(X) \le \underbrace{\log_2 |S_X|}_{\text{uniform}}.$$

**Theorem 2.49.** The expected length $\mathbb{E}[\ell(X)]$ of any uniquely decodable binary code for a random variable $X$ is greater than or equal to the entropy $H(X)$; that is,

$$\mathbb{E}[\ell(X)] \ge H(X)$$

with equality if and only if $2^{-\ell(x)} = p_X(x)$. [5, Thm. 5.3.1]

**Definition 2.50.** Let $L(c, X)$ be the expected codeword length when random variable $X$ is encoded by code $c$.

Let $L^*(X)$ be the minimum possible expected codeword length when random variable $X$ is encoded by a uniquely decodable code $c$:

$$L^*(X) = \min_{\text{UD } c} L(c, X).$$

**2.51.** Given a random variable $X$, let $c_{\text{Huffman}}$ be the Huffman code for this $X$. Then, from the optimality of Huffman code mentioned in 2.37,

$$L^*(X) = L(c_{\text{Huffman}}, X).$$

**Theorem 2.52.** The optimal code for a random variable $X$ has an expected length less than $H(X) + 1$:

$$L^*(X) < H(X) + 1.$$

**2.53.** Combining Theorem 2.49 and Theorem 2.52, we have

Expected length of the optimal UD code
(the same as the expected length of Huffman code)

$$H(X) \leq L^*(X) < H(X) + 1. \tag{3}$$

true for Huffman code

without extension

**Definition 2.54.** Let $L_n^*(X)$ be the minimum expected codeword length per symbol when the random variable $X$ is encoded with $n$-th extension uniquely decodable coding. Of course, this can be achieve by using $n$-th extension Huffman coding.

**2.55.** An extension of (3):

$L_n$

$$H(X) \leq L_n^*(X) < H(X) + \frac{1}{n}. \tag{4}$$

In particular,

$$\lim_{n\to\infty} L_n^*(X) = H(X).$$

In otherwords, by using large block length, we can achieve an expected length per source symbol that is arbitrarily close to the value of the entropy.

**2.56.** Operational meaning of entropy: Entropy of a random variable is the average length of its shortest description.

**2.57. References**

- Section 16.1 in Carlson and Crilly [4]
- Chapters 2 and 5 in Cover and Thomas [5]
- Chapter 4 in Fine [6]
- Chapter 14 in Johnson, Sethares, and Klein [8]
- Section 11.2 in Ziemer and Tranter [18]