x is discrete

X	1	2	3	4	5
$P(x)$	0.1	0.5	0.2	0.1	0.1

$x = 3$

$-\log(p(x=3))$

if we have a distribution like this

The average encoding length will not go below

$$\sum_{i=1}^{5} -p_i \log p_i$$

Entropy

000

001

$\boxed{3}$

800

x	a	b	c	d	e	f	g	h
$p(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$
code	0	10	110	1110	111100	111101	111110	111111

$$-\sum_{i=1}^{8} p_i \log p_i$$

$$\begin{aligned} \mathrm{H}[x] &= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{16}\log_2\frac{1}{16} - \frac{4}{64}\log_2\frac{1}{64} \\ &= 2 \text{ bits} \end{aligned}$$

$$\begin{aligned} \text{average code length} &= \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{16}\times 4 + 4\times\frac{1}{64}\times 6 \\ &= 2 \text{ bits} \end{aligned}$$

$$h(x=1) = -\log p_1$$

$$h(x=2) = -\log p_2$$

$$\sum_{i=1}^{8} -p_i \log p_i = E(h(x))$$

Entropy ↓ expected value of info

Entropy tests the uniformity of a distribution

higher entropy ⇒ more learning, is a very basic understanding

80%

Entropy

x	a	b	c	d	e	f	g	h
$p(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$
code	0	10	110	1110	111100	111101	111110	111111

$$\begin{aligned} H[x] &= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{16}\log_2\frac{1}{16} - \frac{4}{64}\log_2\frac{1}{64} \\ &= 2 \text{ bits} \end{aligned}$$

$$\begin{aligned} \text{average code length} &= \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{16}\times 4 + 4\times\frac{1}{64}\times 6 \\ &= 2 \text{ bits} \end{aligned}$$

if you distribution is non uniform

i.e

X	1	2	3	-----	8
P_X	0.99	0.0001	0.0001		0.002

$$-\sum p_i \log p_i \simeq 0$$

When the distribution is highly uniform

(average information content is a lot)

You need to do a lot of learning!