Information theory

What is information?

- Telecommunications is the transmission of ***information*** over distance to communicate
- So, what is information? How can we define it and measure it for engineering purposes?

Simple example: “SHE-SELLS-SEA-SHELLS”

How much information does it contain?

How many bits do we need to store that information?

ASCII code

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040		Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Source: www.LookupTables.com

Example

- How many bits do we need to encode “SHE-SELLS-SEA-SHELLS”

  → ASCII coding gives 8 bits per character, i.e. 1 byte

  → 20 characters = 20 bytes = 20*8 = 160 bits

- Can we say the information carried by that string is 160 bits?

Compressing information

- Unfortunately is not as easy, for example I could compress that string:
- Define new code based on frequency of letters:

Character	Occurrencies	Relative frequencies	Coding
A	1	0.05	0000
H	2	0.10	0001
-	3	0.15	001
E	4	0.20	10
L	4	0.20	11
S	6	0.30	01

i.e. we use shorter codewords for more frequent characters

➔We can use 49 bits instead of 160…

Why is it interesting?

- This theory is the foundation for:
  - All modern digital communication (obviously)

  But also:
  - Cryptography and cryptanalysis (that's what won World War II)
  - Data compression
  - It has implication in many other fields: physics, linguistics, neurobiology

Definition of information

- So, how can we define and quantify information? – this is a hard question!!
- Refer to Claude Shannon (1916-2001), the father of information theory.

Information is a measure of a reduction of the *entropy* of a random variable

***entropy* is a measure of the uncertainty associated with a random variable**

What???

- Weaver gives an explanation of Shannon's information:

> **Information is a measure of one's freedom of choice in selecting a message. The greater this freedom of choice, the greater the information, the greater is the uncertainty that the message actually selected is some particular one. *Greater freedom of choice, greater uncertainty greater information go hand in hand*.[1]**

- Notice that Shannon does not consider the information carried out by the meaning of a message:

> **"...Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages… "**

[1] C.E.Shannon, W.Weaver, "The Mathematical Theory of Communication"

Shannon entropy

- *Entropy is a measure of the uncertainty associated with a random variable*
- A random variable is used to describe a message in a communication system, it indicates the choice of a message over every possible message

Alice

If Alice sends a message to Bob, she could choose any expression in the English language

Before receiving the message Bob doesn't know what the message is about, it could be anything. At this stage a potential message can be described with a random variable.

Bob

Random variable

- A random variable is a variable whose value is unknown.
- A random variable will follow a certain probability distribution.
- A discrete random variable will have a discrete number of outcomes:
  - flipping a coin has 2 outcomes
  - throwing one die has 6 outcomes

- A continuous random variable will have an infinite number of possible outcomes within a given range:
  - Computer time required to process a certain program
  - The amount of rain that falls in a certain location

Probability distribution

- It defines how likely is that any of the possible values of the random variable would come out.
- Examples:
  - Flip a coin, what is the probability distribution?

    What is the probability to get head or tail?

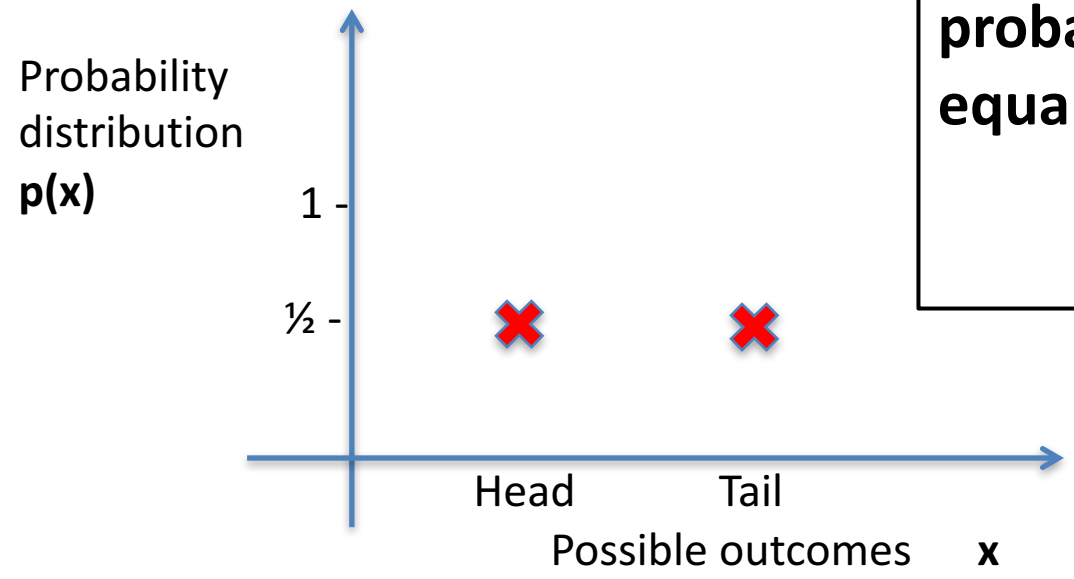
    ➔ There are two possible outcomes, equally likely… so the probability of each one is 1/2 = 0.5.
  - Throw a die, what is the probability distribution?

    ➔ 6 possible outcomes, equally likely… so the probability of each outcome is 1/6 = 0.166666..

Random variable example I

Flipping a coin:

- possible outcomes are head or tail
- Since they are equally likely, the probability distribution is uniform



Notice that the sum of all probabilities needs to be equal to 1:

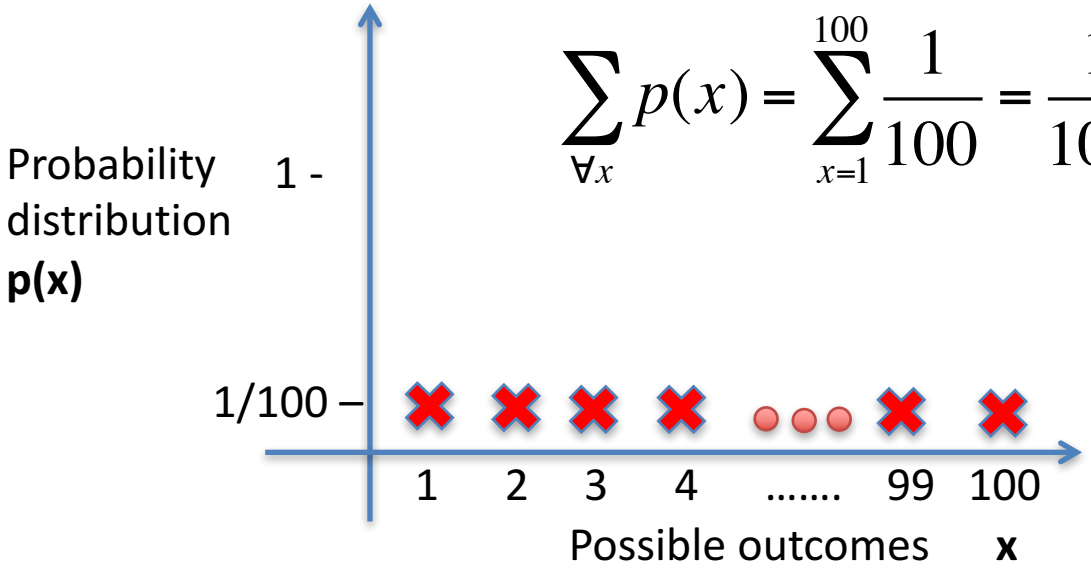
$$\sum_{\forall x} p(x) = 1$$

Random variable example II

Lottery draw:

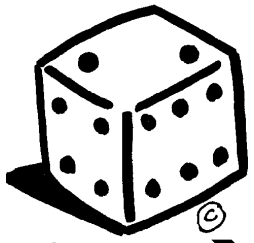
- possible outcomes are numbers, say from 1 to 100
- Since they are equally likely, the probability distribution is uniform

$$\sum_{\forall x} p(x) = \sum_{x=1}^{100} \frac{1}{100} = \frac{1}{100} + \frac{1}{100} + \ldots + \frac{1}{100} = 1$$



Random variable example III

- Throwing one die:



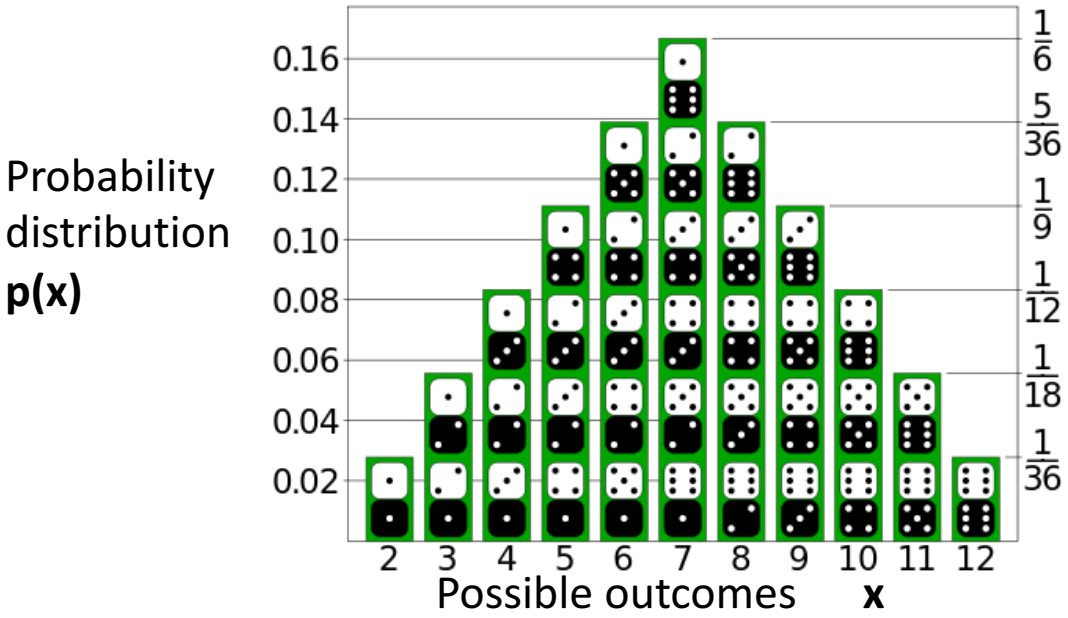
  - Equal outcome for 6 values → uniform distribution with every outcome has probability 1/6

- Throwing two dice?
  - This is the sum of two independent outcomes:
  - What's the probability distribution?

Probability distribution of two dice

- Let's see all possible outcomes, sum them and count the occurrences
- We have 36 possible outcomes
- We check how many times the sum is 1, how many times is 2…

Dice 1	Dice 2	Sum
1	1	2
1	2	3
1	3	4
…	…	…
2	1	3
2	2	4
…	…	…
6	6	12



Shannon entropy

$$H(x) = -\sum_{\forall x} p(x) \log_2 p(x)$$

- H(x) is a measure of the information carried by the random variable and is measured in bits
- *"Greater freedom of choice, greater uncertainty greater information go hand in hand"*
- This means that the higher the uncertainty of the random variable the greater the information it carries

- What carries more information, a random variable with less or more possible outcomes?
- What carries more information, a random variable with uniform or non-uniform distribution?

Information of flipping a coin

$$H(x) = -\sum_{\forall x} p(x)\log_2 p(x) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) =$$

$$= -\left(\frac{1}{2}(-1) + \frac{1}{2}(-1)\right) = -\left(-1\right) = 1$$

- Flipping a coin carries 1 bit of information

Information of flipping an unfair coin

- Say the probability of heads are 0.9 and that of tails 0.1

$$H(x) = -\sum_{\forall x} p(x)\log_2 p(x) = -\left(0.9\log_2 0.9 + 0.1\log_2 0.1\right) =$$

$$= -\left(0.9(-0.152) + 0.1(-3.322)\right) = -\left(-0.469\right) = 0.469$$

Flipping an unfair coin carries less information that one that is totally unpredictable.

- This is understandable: if you know that heads are much more likely to occur, then by telling you that the outcome of the coin was head, you get less information, as you already expected it would probably be head

Information of throwing one die

$$H(x) = -\sum_{\forall x} p(x)\log_2 p(x) = -\left(\frac{1}{6}\log_2\frac{1}{6} + \ldots + \frac{1}{6}\log_2\frac{1}{6}\right) =$$

$$= -6\cdot\left(\frac{1}{6}\log_2\frac{1}{6}\right) = -\log_2\frac{1}{6} = 2.585$$

- Throwing one die carries 2.585 bits of information

A random variable with more possible outcomes carries more information

Uniform distributions carry more information

- Information of throwing two dice

$$H(x) = -\sum_{\forall x} p(x)\log_2 p(x) = -\left(\frac{1}{36}\log_2\frac{1}{36} + \frac{1}{18}\log_2\frac{1}{18} + \ldots\right) =$$

$$= 3.2744$$

- Information of a random distribution between 2 and 12

$$H(x) = -\sum_{\forall x} p(x)\log_2 p(x) = -\sum_{x=2}^{12}\frac{1}{11}\log_2\frac{1}{11} = -\log_2\frac{1}{11} = 3.4594$$

A random variable with uniform distribution carries the maximum amount of information