

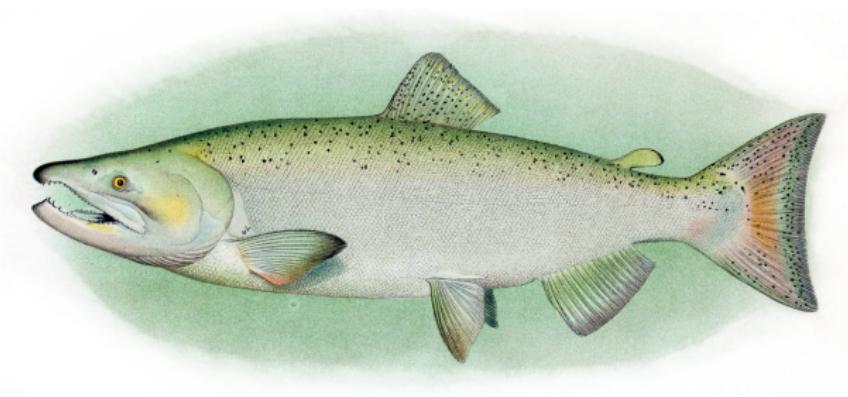
Information Theory

MONDAY

Mathias Winther Madsen

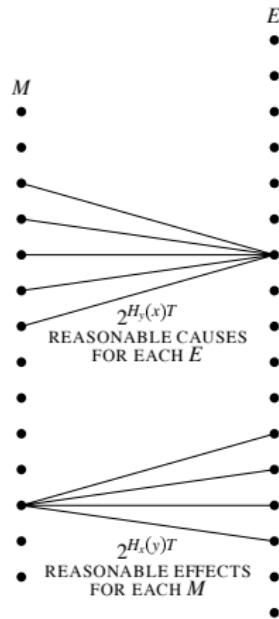
mathias@gmail.com

github.com/mathias-madsen/nasslli2025/



NASSLLI, June 2025

Information Theory



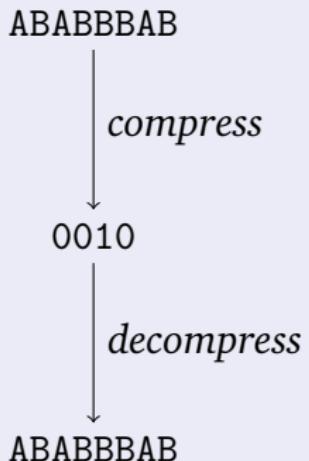
Claude Shannon: “A Mathematical Theory of Communication,”
Bell System Technical Journal, 1948.

Information Theory

THE CHIEF DIFFICULTY ALICE FOUND AT FIRST WAS IN
MANAGING HER FLAMINGO: SHE SUCCEEDED IN GETTING IT
BODY TUCKED AWAY, COMFORTABLY ENOUGH, UNDER HER
ARM, WITH ITS LEGS HANGING DOWN, BUT GENERALLY,
MUST SEE HOW NOT HIS BACK NICELY
TAILORED OUT, AND WAS GOING TO GIVE THE
HEDGEHOG A BASH IN HIS HEAD, IT WOULD WEST
ITSELF ROUND AND LAY IN HIS CROC, WITH SUCH
A ZEE PRIDE THAT NO DOLP
MUST OUGHT: NINE HAD NOT
SEEDED, AND IT IS BEEN NINN,
IT RERED ITSELF IN THE T
EDGEOULD LID SEED, AND A
ROW.

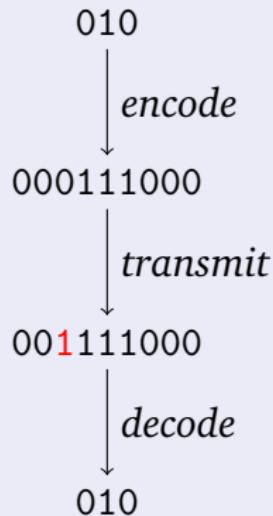
Information Theory

Source Coding



→ source **entropy**

Channel Coding



→ channel **capacity**

Information Theory

- Monday Data compression
- Tuesday Asymptotic equipartition
- Wednesday Random processes
- Thursday Codebreaking
- Friday Error-correcting codes

AS WE SAT OVER OUR VERMOUTHS HE GLORIFIED THE
COMPANY'S BUSINESS, AND BY AND BY I EXPRESSED
CASUALLY MY SURPRISE AT HIM NOT GOING OUT THERE.



```
10100100111110100001110100101011001111000011100  
1001110111110000000011011110111001001110111101  
01100000000110000010100111000100111101101100101  
00011011010110111101010011001111110000010011111  
011101000110101011101101001100111110111000001001  
11000010000000100010101100010100010001110111101  
0011010011111000010011111110100110100111110000  
10011111110101111001101100101101110111011001010  
00100001100111110111010100100000001010100101001  
0011111111101010111111101000000011011011101110  
1101010100001111010110011100010101110101110110  
10001100111101101100001010110101111000000001  
1001111100000100111011001000011
```

Fixed-Width Codes

Jean Maurice
Emilé Baudot:
“Printing Telegraph,”
US Patent 388244,
1888.

ABC \longleftrightarrow 100000011010110

	1	2	3	4	5
A	+	-	-	-	-
B	-	-	+	+	-
C	+	-	+	+	-
D	+	+	+	+	-
E	-	+	-	-	-
F	+	+	-	+	-
G	-	+	+	+	-
H	+	+	-	+	-
I	-	+	+	-	-
J	+	-	-	+	-
K	+	-	-	-	+
L	+	+	-	+	+
M	-	+	-	+	+
N	-	+	+	+	-
O	+	+	+	-	-
P	+	+	+	+	+
Q	+	-	+	+	+
R	-	-	+	+	+
S	-	-	+	-	+
T	+	-	+	-	+
U	+	-	+	-	-
V	+	+	+	-	+
W	+	+	+	-	-
X	-	+	+	-	+
Y	-	+	-	-	-
Z	+	+	-	-	+
�	-	-	-	+	+
�	-	-	-	-	-

Variable-Width Codes

$r = a$	$rlr = f, v$	$rrlr = s$	$lrlr = 3$
$l = e$	$lrr = g$	$rlrr = t$	$llrr = 4$
$rr = i$	$lll = h$	$lrrr = w$	$lllr = 5$
$rl = o$	$llr = l$	$rrll = z$	$llrl = 6$
$lr = u$	$lr l = m$	$rlrl = o$	$lr ll = 7$
$ll = b$	$rll = n$	$rl lr = 1$	$rlli = 8$
$rrr = c, k$	$rrrr = p$	$lrrl = 2$	$llll = 9$
$rrl = d$	$rrrl = r$		

HELLO \longleftrightarrow LLL L LLR LLR RL

The Gauss-Weber code; from John Jacob Fahie:
A History of Electric Telegraphy, to the Year 1837, 1884.

Variable-Width Codes

A	B	C
0_-	1_-	00_-

ABACB_B \longleftrightarrow 0_1_00_1_1_-

A	B	C
0	1	-

ABACB_B \longleftrightarrow 010_11

Unique Decodability

Definition

A set of codewords is **uniquely decodable** if no two sequences of codewords produce identical texts.

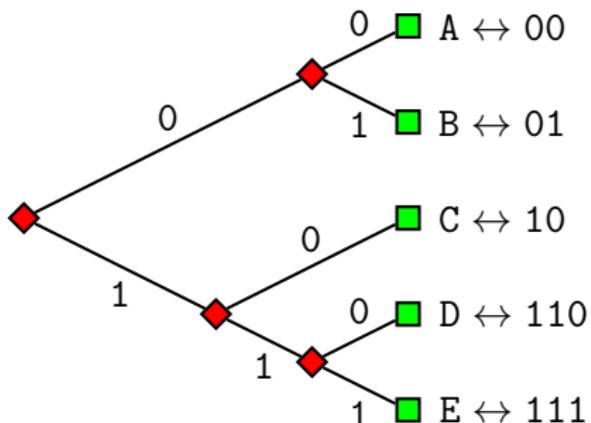
Exercises

1. $\{00, 0, 1\}$
2. $\{0, 10, 11\}$
3. $\{0, 01, 11\}$
4. $\{01, 10, 1001\}$
5. $\{00, 1100, 1111\}$
6. $\{02, 12, 002, 012, 102, 112\}$

Prefix Codes

Definition

A code is a **prefix code** if no codeword is a prefix of another.

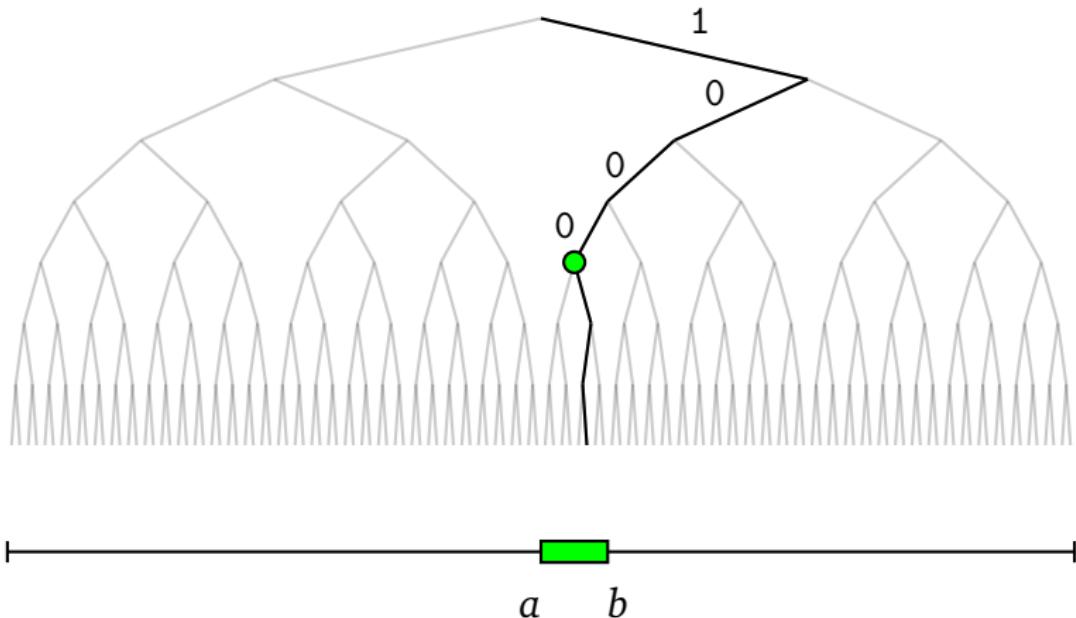


Exercises

1. $\{0, 10, 11\}$
2. $\{0, 1, 11\}$
3. $\{0, 1\}$
4. $\{02, 12, 012\}$
5. $\{0, 01\}$

ABEADCB \longleftrightarrow 0001111001101001

Strings ~ Paths ~ Intervals



$$a = s_1/2 + s_2/4 + s_3/8 + \cdots + s_w/2^w$$

$$b = a + 1/2^w$$

Kraft-McMillan

Theorem: Kraft-McMillan Theorems

There is a uniquely decodable binary code with codeword lengths w_1, w_2, \dots, w_k if and only if

$$\underbrace{\left(\frac{1}{2}\right)^{w_1} + \left(\frac{1}{2}\right)^{w_2} + \cdots + \left(\frac{1}{2}\right)^{w_k}}_{\text{code "footprint"} } \leq 1$$

Exercises

1. $\{0, 1\}$
2. $\{0, 01\}$
3. $\{0, 1, 11\}$
4. $\{0, 10, 11\}$
5. $\{0, 0\}$

Leon G. Kraft: *A Device for Quantizing, Grouping, and Coding Amplitude Modulated Pulses*, MIT master's thesis, 1949.

Brockway McMillan: “Two Inequalities Implied by Unique Decipherability,” *IEEE Transactions on Information Theory*, 1956.

Kraft's Inequality

Kraft Codes

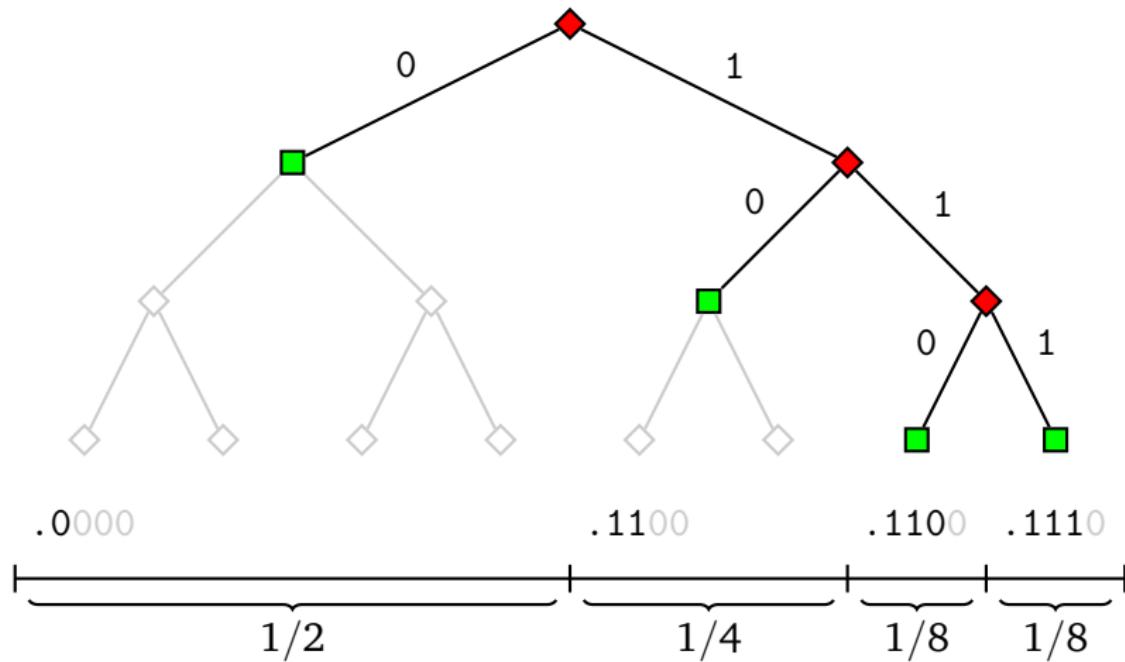
Pick codewords of length $w_1 \geq w_2 \geq \dots \geq w_k$, always using the alphabetically first codeword not blocked by a previous codeword.

Exercises

1. $1 \geq 2 \geq 2$
2. $2 \geq 2 \geq 3 \geq 3 \geq 3 \geq 4 \geq 4$
3. $2 \geq 2 \geq 3 \geq 3 \geq 3 \geq 3$

Kraft's Inequality

First available codewords of lengths $1 \leq 2 \leq 3 \leq 3$:



McMillan's Theorem: Small Footprints

Multiword sentences

Let $C^{\otimes n}$ be the set of all sentences of n codewords from C .

$$\{0, 10\}^{\otimes 1} = \{0, 01, 10\}$$

$$\{0, 10\}^{\otimes 2} = \{00, 010, 100, 1010\}$$

$$\begin{aligned}\{0, 10\}^{\otimes 3} = & \{000, 0010, 0100, 01010, \\ & 1000, 10010, 10100, 101010, \\ & 1100, 11010, 11100, 111010\}\end{aligned}$$

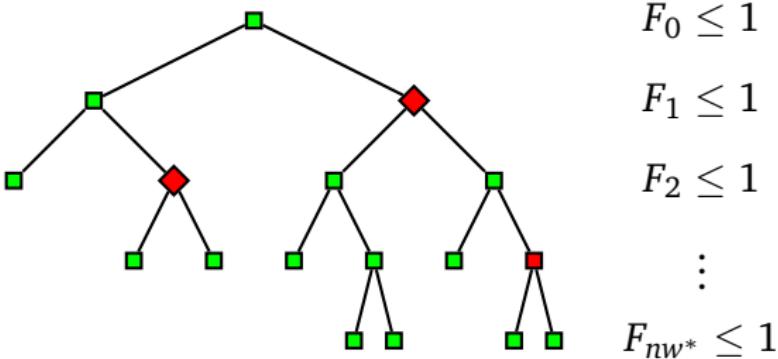
⋮

McMillan's Theorem: Small Footprints

Lemma: Linear Upper Bound

If C is uniquely decodable with $w_1, w_2, \dots, w_k \leq w^*$, then

$$F(C^{\otimes n}) \leq 1 + nw^*$$



McMillan's Theorem: Large Footprints

Lemma: Sentence Footprint Exponential in Code Footprint

If C is uniquely decodable, then

$$F(C^{\otimes n}) = F(C)^n$$

Proof: Listing sentences \sim combining codewords

$$\sum_{i=1}^k \sum_{j=1}^k 2^{-(w_i + w_j)} = \left(\sum_{i=1}^k 2^{-w_i} \right) \times \left(\sum_{j=1}^k 2^{-w_j} \right)$$

Least Average Codeword Length

With letter probabilities p_1, p_2, \dots, p_k :



$$\begin{aligned} & \text{minimize } \sum_{i=1}^k p_i w_i \\ & \text{subject to } \sum_{i=1}^k 2^{-w_i} \leq 1 \end{aligned}$$

Least Average Codeword Length

Solution:

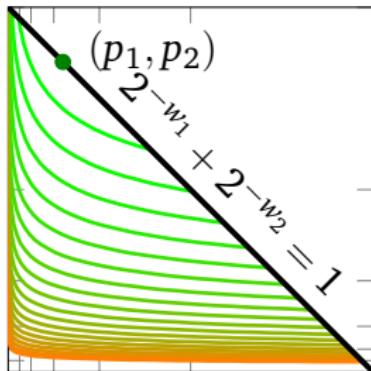
$$w_i^* = \log_2 \frac{1}{p_i}$$

$$\sum_{i=1}^k p_i \log_2 \frac{1}{p_i}$$

$$\sum_{i=1}^k p_i w_i$$

$$2^{-w_2}$$

$$2^{-w_1}$$



Entropy

Definition

$$H = \sum_{i=1}^k p_i \log_2 \frac{1}{p_i}$$

Claude Shannon: “A Mathematical Theory of Communication,”
Bell System Technical Journal, 1948.

Entropy

Definition

$$H = \sum_{i=1}^k p_i \log_2 \frac{1}{p_i}.$$

1.

x	A	B
$p(x)$	1/2	1/2

2.

x	A	B	C	D
$p(x)$	1/2	1/4	1/8	1/8

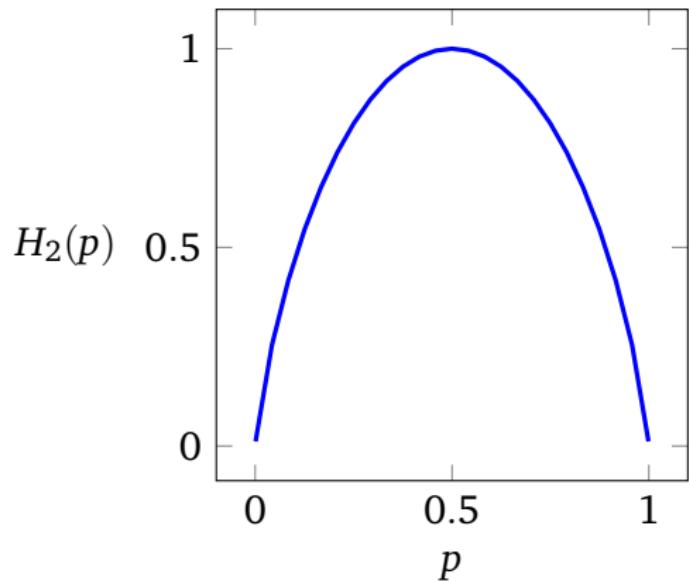
3.

x	A	B	C
$p(x)$	1/3	1/3	1/3

4.

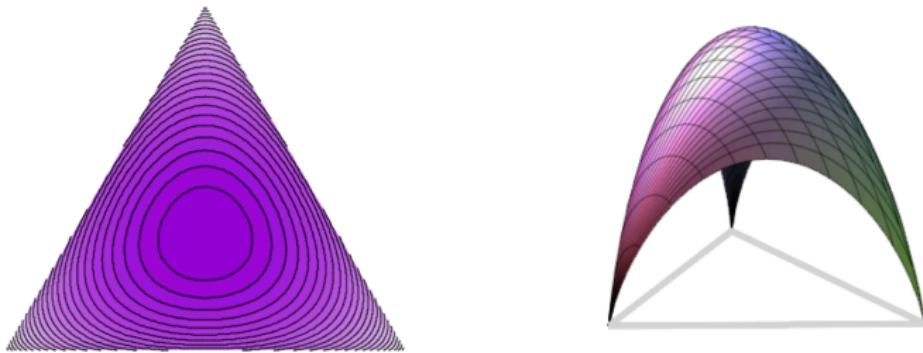
x	A	B	C
$p(x)$	1/2	1/3	1/6

Binary Entropy



p_1	p_2	H
0.0	1.0	0.00
0.1	0.9	0.47
0.2	0.8	0.72
0.3	0.7	0.88
0.4	0.6	0.97
0.5	0.5	1.00

Tertiary Entropy



p_1	p_2	p_3	H
1/3	1/3	1/3	1.58
1/2	1/2	0	1.00
1/2	1/4	1/4	1.50

Lower and Upper bounds

Theorem

The least average codeword length L^* for a random letter with entropy H satisfies

$$H \leq L^* \leq H + 1$$

Proof. The codeword lengths

$$w_i = \left\lceil \log \frac{1}{p_i} \right\rceil \leq \log_2 \frac{1}{p_i} + 1$$

satisfy Kraft's inequality.

Huffman Coding

A	B	C	D	E
.35	.25	.20	.15	.05

Huffman Coding

x	$p(x)$	$-\log_2 p(x)$	w	codeword
A	.0634	3.98	4	1001
B	.0135	6.21	6	011101
C	.0242	5.37	5	00011
D	.0321	4.96	5	10100
E	.0980	3.35	3	001
F	.0174	5.84	6	101111
G	.0165	5.92	6	101011
H	.0438	4.51	5	11011
I	.0552	4.18	4	0110
J	.0009	10.17	9	011100000
K	.0061	7.35	7	0111001
L	.0336	4.89	5	10110
M	.0174	5.85	6	101110
N	.0551	4.18	4	0101
O	.0622	4.01	4	1000
P	.0180	5.80	6	110100

x	$p(x)$	$-\log_2 p(x)$	w	codeword
Q	.0008	10.33	10	0111000100
R	.0470	4.41	4	0000
S	.0502	4.32	4	0100
T	.0729	3.78	4	1100
U	.0234	5.42	5	00010
V	.0075	7.06	7	011110
W	.0156	6.00	6	011110
X	.0014	9.46	9	011100001
Y	.0160	5.97	6	101010
Z	.0005	11.04	11	011100010111
¶	.0084	6.89	7	0111111
-	.1741	2.52	3	111
,	.0019	9.06	9	011100011
,	.0117	6.42	7	1101011
.	.0109	6.52	7	1101010
?	.0003	11.56	11	01110001010

Exercises

1. construct a Huffman code for

A	B	C	D
.1	.2	.3	.4

2. encode the message DCDDAB according to your code
3. decode the message 0010100001 according to your code
4. compute the expected codeword length of your code
5. compute the entropy of the original distribution