

# Lecture 1

Introduction to lossless compression

EE 274: Data Compression - Lecture 1

Plan: Lecture 1-3: theory and concepts from information theory

### A simple probability distribution

#### Consider:

- Alphabet  $\mathcal{X} = \{A, B, C, D\}$
- ullet Uniform probability distribution:  $P(A)=P(B)=P(C)=P(D)=rac{1}{4}$

A text file generating by independently sampling one million symbols from this distribution:

```
$ cat abcd.txt
ACABDADCBDDC....
```

What is the size of this file?

### Bits and bytes

- bit: a unit of information expressed as either a 0 or 1 in binary notation.
- byte: a group of eight bits operated on as a unit.

1 byte (B) = 8 bits

1 kilobyte (KB) = 1000 bytes = 8000 bits

So on for MB, GB, TB, PB, EB, ...

**Note:** Sometimes we like to use powers of two, e.g., 1 kilobyte = 1024 bytes.

# abcd.txt

Size on disk: 1 MB (1 million bytes).

Why 1 byte per letter/character?

# **ASCII TABLE**

| 2 2 [START OF TEXT] 34 22 " 66 42 B 98 62 b 33 3 [END OF TEXT] 35 23 # 67 43 C 99 663 C b 44 4 [END OF TRANSMISSION] 36 24 \$ 68 44 D 100 64 d 65 5 [ENQUIRY] 37 25 % 69 45 E 101 65 e 66 6 [ACKNOWLEDGE] 38 26 & 70 46 F 102 666 F 7 7 [BELL] 39 27 ' 71 47 G 103 67 9 8 8 [BACKSPACE] 40 28 ( 72 48 H 104 68 h 104 68 h 105 69 9 [HORIZONTAL TAB] 41 29 ) 73 49 I 105 69 i 10 A [LINE FEED] 42 2A * 74 4A J 106 6A j 11 B [VERTICAL TAB] 43 2B + 75 4B K 107 6B k 12 C [FORM FEED] 44 2C , 76 4C L 108 6C I 13 D [CARRIAGE RETURN] 45 2D - 77 4D M 109 6D m 14 E [SHIFT OUT] 46 2E . 78 4E N 110 6E m 15 F [SHIFT IN] 47 2F / 79 4F 0 111 6F 0 111 6F 0 111 6F 0 10 [DATA LINK ESCAPE] 48 30 0 8 50 P 112 70 P 117 11 [DEVICE CONTROL 1] 49 31 1 81 51 Q 113 71 G 113 71 G 114 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 77 T 10 11 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 77 T 10 11 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 77 T 10 11 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 77 T 10 11 [DEVICE CONTROL 2] 50 32 2 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8  | Decimal  | Hex | Char                   | Decimal | Нех | Char    | <sub>I</sub> Decimal | Нех | Char | <sub>I</sub> Decimal | Нех | Char  |
|--|----------|-----|------------------------|---------|-----|---------|----------------------|-----|------|----------------------|-----|-------|
| 2 2 (START OF TEXT) 34 22 " 666 42 B 98 62 b 3 3 3 (END OF TEXT) 35 23 # 67 43 C 99 63 C 4 4 4 (END OF TRANSISSION) 36 24 \$ 68 44 D 100 64 C 5 5 [ENQUIRY] 37 25 % 669 45 E 101 65 e 6 6 6 [ACKNOWLEDGE] 38 26 & 70 46 F 102 66 f 7 7 7 [BELL] 39 27 ' 71 47 G 103 67 8 8 8 [BACKSPACE] 40 28 ( 72 48 H 104 68 h 9 9 (HORIZONTAL TAB) 41 29 ) 73 49 I 105 69 i 10 A [LINE FEED] 42 2A * 74 4A J 106 6A j 11 B [VERTICAL TAB] 43 2B + 75 4B K 107 6B k 12 C [FORM FEED] 44 2C , 76 4C L 108 6C I 13 D [CARRIAGE RETURN] 45 2D - 77 4D M 109 6D m 14 E [SHIFT OUT] 46 2E . 78 4E N 110 6E 15 F [SHIFT IN] 47 2F / 79 4F O 111 6F 0 16 10 [DATA LINK ESCAPE] 48 30 0 80 50 P 112 70 P 17 11 [DEVICE CONTROL 1] 49 31 1 81 51 Q 113 71 Q 17 11 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T 19 13 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 20 14 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 20 14 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 20 14 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 21 15 [NEGATIVE ACKNOWLEDGE] 53 35 5 85 55 U 117 75 U 22 16 [SYNCHRONOUS IDLE] 54 36 6 86 56 V 118 76 V 23 17 [END OF TRANS. BLOCK] 55 37 7 87 57 W 119 77 W 24 18 [CANCEL] 58 38 8 88 88 X 120 78 25 19 [END OF MEDIUM] 57 39 9 89 59 Y 121 79 Y 26 1A [SUBSTITUTE] 58 3A : 90 5A Z 122 7A Z 27 1B [ESCAPE] 59 3B ; 91 5B [ 123 7B { 28 1C [FILE SEPARATOR] 61 3D = 93 5D ] 125 7D }  | )        | 0   | [NULL]                 | 32      | 20  | [SPACE] | 64                   | 40  | @    | 96                   | 60  | *     |
| 3 3 [END OF TEXT] 34 22 66 42 8 96 62 4  4 4 [END OF TRANSMISSION] 36 24 \$ 68 44 D 100 64 d  5 5 [ENQUIRY] 37 25 % 69 45 E 101 65 66 6 6 [ACKNOWLEDGE] 38 26 & 70 46 F 102 66 f  7 7 7 [BELL] 39 27 ' 71 47 G 103 67 9  8 8 8 [BACKSPACE] 40 28 ( 72 48 H 104 68 H 104 68 H 105 69 i  10 A [LINE FEED] 42 2A * 74 4A J 106 6A J 105 69 i  11 B [VERTICAL TAB] 43 2B + 75 4B K 107 6B k 1  12 C [FORM FEED] 44 2C , 76 4C L 108 6CC I  13 D [CARRIAGE RETURN] 45 2D - 77 4D M 109 6D m  14 E [SHIFT OUT] 46 2E . 78 4E N 110 6E m  15 F [SHIFT OUT] 46 2E . 78 4E N 110 6E m  16 10 [DATA LINK ESCAPE] 48 30 0 80 50 P 112 70 p  17 11 [DEVICE CONTROL 1] 49 31 1 81 51 Q 113 71 9  18 12 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 T  19 13 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S  20 14 [DEVICE CONTROL 3] 51 33 35 5 85 55 U 117 75 Q  21 15 [NEGATIVE ACKNOWLEDGE] 53 35 5 85 55 U 117 75 Q  22 16 [SYNCHRONOUS IDLE] 54 36 6 88 6 56 V 118 76 V 122 77 Q  24 18 [CANCEL] 56 38 8 8 88 58 X 120 78 X 122 78 Z  27 18 [ESCAPE] 59 38 ; 91 58 [123 78 [22 74 22 74 22 75 ]  28 1C [FILE SEPARATOR] 61 3D = 93 5D ] 125 7D 3  | L        | 1   | [START OF HEADING]     | 33      | 21  | 1       | 65                   | 41  | Α    | 97                   | 61  | a     |
| 4  | 2        | 2   | [START OF TEXT]        | 34      | 22  | "       | 66                   | 42  | В    | 98                   | 62  | b     |
| 5  | 3        | 3   | [END OF TEXT]          | 35      | 23  | #       | 67                   | 43  | С    | 99                   | 63  | c     |
| 6 6 [ACKNOWLEDGE] 38 26 & 70 46 F 102 66 f 7 7 7 [BELL] 39 27 ' 71 47 G 103 67 9 8 8 8 [BACKSPACE] 40 28 ( 72 48 H 104 68 h 105 69 i 10 A [LINE FEED] 42 2A * 74 4A J 106 6A j 11 B [VERTICAL TAB] 41 29 ) 73 49 i 105 69 i 10 A [LINE FEED] 42 2A * 74 4A J 106 6A j 11 B [VERTICAL TAB] 43 2B + 75 4B K 107 6B k 12 C [FORM FEED] 44 2C , 76 4C L 108 6C I 13 D [CARRIAGE RETURN] 45 2D - 77 4D M 109 6D n 14 E [SHIFT OUT] 46 2E . 78 4E N 110 6E n 15 F [SHIFT IN] 47 2F / 79 4F O 111 6F O 111 6F O 111 6F O 111 1 [DEVICE CONTROL 1] 49 31 1 81 51 Q 113 71 Q 113 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 r 19 13 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 12 C 14 [DEVICE CONTROL 4] 52 34 4 84 54 T 116 74 t 11 [DEVICE CONTROL 4] 54 T 11 [DEV | 1        | 4   | [END OF TRANSMISSION]  | 36      | 24  | \$      | 68                   | 44  | D    | 100                  | 64  | d     |
| 7  | 5        | 5   | [ENQUIRY]              | 37      | 25  | %       | 69                   | 45  | E    | 101                  | 65  | e     |
| 8 8 [BACKSPACE] 40 28 ( 72 48 H 104 68 h 9 9 [HORIZONTAL TAB] 41 29 ) 73 49 I 105 69 i 10 A [LINE FEED] 42 2A * 74 4A J 106 6A j 11 B [VERTICAL TAB] 43 2B + 75 4B K 107 6B k 12 C [FORM FEED] 44 2C , 76 4C L 108 6C I 13 D [CARRIAGE RETURN] 45 2D - 77 4D M 109 6D m 14 E [SHIFT OUT] 46 2E . 78 4E N 110 6E m 15 F [SHIFT IN] 47 2F / 79 4F O 111 6F O I I 111 6F O I I I I I I I I I I I I I I I I I I  | 5        | 6   | [ACKNOWLEDGE]          | 38      | 26  | &       | 70                   | 46  | F    | 102                  | 66  | f     |
| 9 9 [HORIZONTÁL TAB] 41 29 ) 73 49 I 105 69 i 10 A [LINE FEED] 42 2A * 74 4A J 106 6A j 11 B [VERTICAL TAB] 43 2B + 75 4B K 107 6B k 12 C [FORM FEED] 44 2C , 76 4C L 108 6C I 13 D [CARRIAGE RETURN] 45 2D - 77 4D M 109 6D n 14 E [SHIFT OUT] 46 2E . 78 4E N 110 6E n 15 F [SHIFT IN] 47 2F / 79 4F O 111 6F o 16 10 [DATA LINK ESCAPE] 48 30 0 80 50 P 112 70 p 17 11 [DEVICE CONTROL 1] 49 31 1 81 51 Q 113 71 1 18 12 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 r 19 13 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 20 14 [DEVICE CONTROL 4] 52 34 4 84 54 T 116 74 t 21 15 [NEGATIVE ACKNOWLEDGE] 53 35 5 85 55 U 117 75 U 22 16 [SYNCHRONOUS IDLE] 54 36 6 86 56 V 118 76 23 17 [END OF TRANS. BLOCK] 55 37 7 87 57 W 119 77 W 24 18 [CANCEL] 56 38 8 88 88 58 X 120 78 X 25 19 [END OF MEDIUM] 57 39 9 89 59 Y 121 79 Y 26 1A [SUBSTITUTE] 58 3A : 90 5A Z 122 7A Z 27 1B [ESCAPE] 59 3B ; 91 5B [ 123 7B [ 28 1C [FILE SEPARATOR] 61 3D = 93 5D ] 125 7D   | <i>†</i> | 7   | [BELL]                 | 39      | 27  | 1       | 71                   | 47  | G    | 103                  | 67  | g     |
| 10   |          | 8   | [BACKSPACE]            | 40      | 28  | (       | 72                   | 48  | H    | 104                  | 68  | h     |
| 11 B [VERTICAL TAB]  | )        | 9   | (HORIZONTAL TAB)       | 41      | 29  | )       | 73                   | 49  | 1    | 105                  | 69  | i i   |
| 12   | 10       | Α   | [LINE FEED]            | 42      | 2A  | *       | 74                   | 4A  | J    | 106                  | 6A  | j     |
| 13 D [CARRIAGE RETURN]   | 11       | В   | [VERTICAL TAB]         | 43      | 2B  | +       | 75                   | 4B  | K    | 107                  | 6B  | k     |
| 14 E [SHIFT OUT]   | 12       | C   | (FORM FEED)            | 44      | 2C  | ,       | 76                   | 4C  | L    | 108                  | 6C  | 1     |
| 15 F [SHIFT IN] 47 2F / 79 4F 0 111 6F 0 16 10 [DATA LINK ESCAPE] 48 30 0 80 50 P 112 70 p 17 11 [DEVICE CONTROL 1] 49 31 1 81 51 Q 113 71 p 18 12 [DEVICE CONTROL 2] 50 32 2 82 52 R 114 72 r 19 13 [DEVICE CONTROL 3] 51 33 3 83 53 S 115 73 S 120 14 [DEVICE CONTROL 4] 52 34 4 84 54 T 116 74 t 17 116 74 t 18 115 [NEGATIVE ACKNOWLEDGE] 53 35 5 85 55 U 117 75 u 18 17 [END OF TRANS. BLOCK] 55 37 7 87 57 W 119 77 w 19 19 [END OF TRANS. BLOCK] 55 38 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8  | 13       | D   | [CARRIAGE RETURN]      | 45      | 2D  | -       | 77                   | 4D  | M    | 109                  | 6D  | m     |
| 16       10       [DATA LINK ESCAPE]       48       30       0       80       50       P       112       70       p         17       11       [DEVICE CONTROL 1]       49       31       1       81       51       Q       113       71       q         18       12       [DEVICE CONTROL 2]       50       32       2       82       52       R       114       72       r         19       13       [DEVICE CONTROL 3]       51       33       3       83       53       S       115       73       s         20       14       [DEVICE CONTROL 4]       52       34       4       84       54       T       116       74       t         21       15       [NEGATIVE ACKNOWLEDGE]       53       35       5       85       55       U       117       75       u         22       16       [SYNCHRONOUS IDLE]       54       36       6       86       56       V       118       76       v         23       17       [END OF TRANS. BLOCK]       55       37       7       87       57       W       119       77       w         24 <t< td=""><td>14</td><td>E</td><td>[SHIFT OUT]</td><td>46</td><td>2E</td><td></td><td>78</td><td>4E</td><td>N</td><td>110</td><td>6E</td><td>n</td></t<>   | 14       | E   | [SHIFT OUT]            | 46      | 2E  |         | 78                   | 4E  | N    | 110                  | 6E  | n     |
| 17   | 15       | F   | [SHIFT IN]             | 47      | 2F  | 1       | 79                   | 4F  | 0    | 111                  | 6F  | 0     |
| 18   | 16       | 10  | [DATA LINK ESCAPE]     | 48      | 30  | 0       | 80                   | 50  | P    | 112                  | 70  | р     |
| 19   | 17       | 11  | [DEVICE CONTROL 1]     | 49      | 31  | 1       | 81                   | 51  | Q    | 113                  | 71  | q     |
| 20   | 18       | 12  | [DEVICE CONTROL 2]     | 50      | 32  | 2       | 82                   | 52  | R    | 114                  | 72  | r     |
| 21   | 19       | 13  | [DEVICE CONTROL 3]     | 51      | 33  | 3       | 83                   | 53  | S    | 115                  | 73  | S     |
| 22 16 [SYNCHRONOUS IDLE] 54 36 6 86 56 V 118 76 V 23 17 [END OF TRANS. BLOCK] 55 37 7 87 57 W 119 77 W 124 18 [CANCEL] 56 38 8 88 58 X 120 78 X 25 19 [END OF MEDIUM] 57 39 9 89 59 Y 121 79 Y 26 1A [SUBSTITUTE] 58 3A : 90 5A Z 122 7A Z 27 1B [ESCAPE] 59 3B ; 91 5B [ 123 7B { 28 1C [FILE SEPARATOR] 60 3C < 92 5C \ 1D [GROUP SEPARATOR] 61 3D = 93 5D ] 125 7D }  | 20       | 14  | [DEVICE CONTROL 4]     | 52      | 34  | 4       | 84                   | 54  | T    | 116                  | 74  | t     |
| 23   | 21       | 15  | [NEGATIVE ACKNOWLEDGE] | 53      | 35  | 5       | 85                   | 55  | U    | 117                  | 75  | u     |
| 24     18     [CANCEL]     56     38     8     88     58     X     120     78     x       25     19     [END OF MEDIUM]     57     39     9     89     59     Y     121     79     y       26     1A     [SUBSTITUTE]     58     3A     :     90     5A     Z     122     7A     z       27     1B     [ESCAPE]     59     3B     ;     91     5B     [     123     7B     {       28     1C     [FILE SEPARATOR]     60     3C     <  | 22       | 16  | [SYNCHRONOUS IDLE]     | 54      | 36  | 6       | 86                   | 56  | V    | 118                  | 76  | v     |
| 25     19 [END OF MEDIUM]     57     39     9     89     59     Y     121     79     Y       26     1A [SUBSTITUTE]     58     3A     :     90     5A     Z     122     7A     z       27     1B [ESCAPE]     59     3B     ;     91     5B     [     123     7B     {       28     1C [FILE SEPARATOR]     60     3C     <  | 23       | 17  | [END OF TRANS. BLOCK]  | 55      | 37  | 7       | 87                   | 57  | w    | 119                  | 77  | w     |
| 26   | 24       | 18  | [CANCEL]               | 56      | 38  | 8       | 88                   | 58  | X    | 120                  | 78  | x     |
| 27   | 25       | 19  | [END OF MEDIUM]        | 57      | 39  | 9       | 89                   | 59  | Υ    | 121                  | 79  | у     |
| 28 1C [FILE SEPARATOR] 60 3C < 92 5C \ 124 7C   29 1D [GROUP SEPARATOR] 61 3D = 93 5D ] 125 7D }   |          | 1A  | (SUBSTITUTE)           |         | 3A  | 1       | 90                   | 5A  | Z    | 122                  | 7A  | z     |
| 29 1D [GROUP SEPARATOR] 61 3D = 93 5D ] 125 7D }   | 27       | 1B  | [ESCAPE]               |         |     | ;       | 91                   | 5B  | [    |                      | 7B  | {     |
|  | 28       | 1C  | [FILE SEPARATOR]       | 60      | 3C  | <       | 92                   | 5C  | \    | 124                  | 7C  | 1     |
| 30 1F (RECORD SEPARATOR)   62 3E >   94 5F ^   126 7F  | 29       | 1D  | [GROUP SEPARATOR]      | 61      | 3D  | =       | 93                   | 5D  | 1    | 125                  | 7D  | }     |
|  | 30       | 1E  | [RECORD SEPARATOR]     | 62      | 3E  | >       | 94                   | 5E  | ^    | 126                  | 7E  | ~     |
| 31 1F [UNIT SEPARATOR] 63 3F ? 95 5F _ 127 7F [I   | 31       | 1F  | [UNIT SEPARATOR]       | 63      | 3F  | ?       | 95                   | 5F  | _    | 127                  | 7F  | [DEL] |

EE 274: Data Compression - Lecture 1

### **ASCII Table**

| Symbol | ASCII code |
|--------|------------|
| Α      | 1000001    |
| В      | 1000010    |
| С      | 1000011    |
| D      | 1000100    |

8 bits = 1 byte per symbol.

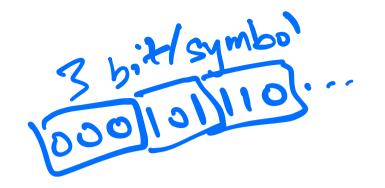
Can we do better?

### Fixed bitwidth code

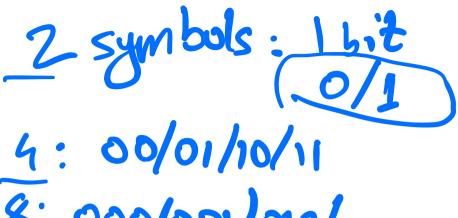
| Symbol | Code |
|--------|------|
| A      | 00   |
| В      | 01   |
| С      | 10   |
| D      | 11   |

Bits/symbol?

Decoding?



# Fixed bitwidth code



8:000/001/010/...

 $k = |\mathcal{S}|$  different symbols implies at least  $\lceil \log_2 k \rceil$  bits per symbol in a fixed bitwidth code.

Can we do better? In the uniform distribution example above?

### **Uniform distribution**

| Symbol | Probability |  |  |
|--------|-------------|--|--|
| A      | 0.5         |  |  |
| В      | 0.5         |  |  |

Fixed bitwidth code: 1 bit/symbol

### Non-uniform distribution

| Symbol | Probability |
|--------|-------------|
| Α      | 0.49        |
| В      | 0.49        |
| С      | 0.01        |
| D      | 0.01        |

Fixed bitwidth code: 2 bits/symbol

Can we do better? Closer to the previous page's 1 bit/base?

### Non-uniform distribution

| Symbol | Probability |
|--------|-------------|
| Α      | 0.49        |
| В      | 0.49        |
| С      | 0.01        |
| D      | 0.01        |

**Solution 1:** C and D are low probability, let's just lose them - Lossy Compression (not commonly used for text/database/log data).

EE 274: Data Compression - Lecture 1

#### Non-uniform distribution

| Symbol | Probability |
|--------|-------------|
| A      | 0.49        |
| В      | 0.49        |
| С      | 0.01        |
| D      | 0.01        |

Solution 2: Variable length codes: Use fewer bits for more probable symbols.

# Variable length codes

Use fewer bits for more probable symbols

| Symbol | Probability | Code  | 0110111                   |
|--------|-------------|-------|---------------------------|
| A'     | 0.49        | 0 <   | 1 bit 2 105 bits/symbol   |
| В.     | 0.49        | 10 '  | -25its - 105 6rts/3941901 |
| С      | 0.01        | 110   |                           |
| D      | 0.01        | 111 - |                           |

How to evaluate coding efficiency? Expected code length.

14

# **Expected code length**

"Compressed size/Uncompressed size" - often in units bits/symbol.

Also sometimes called compression rate/compression ratio

Warning: There's some variability in notation and definitions of these terms so be careful.

Let l(x) denote the code length for symbol x with probability P(x), where  $x \in \mathcal{X}$ .

Expected code length:  $\mathbb{E}[l(X)] = \sum_{x \in \mathcal{X}} P(x) l(x)$ 

### **Expected code length**

| Symbol | Probability | Code | T(x) |
|--------|-------------|------|------|
| Α      | 0.49        | 0    |      |
| В      | 0.49        | 10   | 2    |
| С      | 0.01        | 110  | 3    |
| D      | 0.01        | 111  | 3    |

1-53

Expected code length: 
$$\mathbb{E}[l(X)] = ?$$
  $P(A) \times L(A) + P(B) \wedge L(B)$ 

# **Expected code length**

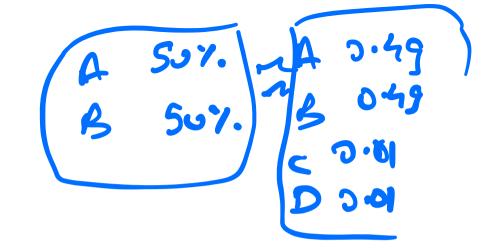
| Ø: | 11111010000 | 111110 |
|----|-------------|--------|
|    | Donado 1    |        |

| Symbol | Probability | Code | l(x) |
|--------|-------------|------|------|
| Α      | 0.49        | 0    | 1    |
| В      | 0.49        | 10   | 2    |
| С      | 0.01        | 110  | 3    |
| D      | 0.01        | 111  | 3    |

$$\mathbb{E}[l(X)] = 0.49 imes 1 + 0.49 imes 2 + 0.01 imes 3 + 0.01 imes 3 = 1.53$$
 bits/symbol

# Thoughts and conclusion

• Is the code above lossless? Can you decode it? <- homework for next lecture!



# Thoughts and conclusion

- Is the code above lossless? Can you decode it? <- homework for next lecture!</li>
- The non-uniform distribution above seems "worse" but "similar" to the uniform distribution on just A and B.

EE 274: Data Compression - Lecture 1

# Thoughts and conclusion

- Is the code above lossless? Can you decode it? <- homework for next lecture!</li>
- The non-uniform distribution above seems "worse" but "similar" to the uniform distribution on just A and B.
- In the next few lectures, we will learn how to compute the optimal compression rate and how we can get close to 1.14 bits/symbol for the above distribution (and no better).

# Thank you!