Defs:

- "expected code word length" for a given code book C

$$L := \mathbb{E}[l(x)] = \sum_{x \in \mathcal{X}} p(x)\, l(x)$$

($p(x)$: prob. to find x; $l(x)$: length of $C(x)$)

- A code C defines a "uniquely decodable" symbol code iff C^* is injective, i.e.: (property of C^*)

$$\forall \underline{x}, \underline{x}' \in \mathcal{X}^* \text{ with } \underline{x} \neq \underline{x}': \quad C^*(\underline{x}) \neq C^*(\underline{x}')$$

- C is a "prefix-free symbol code" (= "prefix code") iff: (property of C)

$\forall x, x' \in \mathcal{X}$ with $x \neq x'$: $C(x)$ is not a prefix of $C(x')$ (i.e. $C(x')$ does not start with $C(x)$)

Claim: C is a prefix code $\Rightarrow$ C is uniquely decodable (but inverse in general not true)

New example:

x	$C(x)$	$C'(x)$	$p(x)$
"a"	00	0 ← shorter by 1 bit with prd. 0.4	0.4
"b"	01	10	0.3
"c"	10	110 } longer by 1 bit with prd. 0.3	0.2
"d"	11	111 }	0.1
L	2	1.9	

THEORETICAL BOUNDS FOR LOSSLESS COMPRESSION IV

27 April 2021

Questions we will answer today:
- ↳ theoretical bound on bitrate of lossless compression
- ↳ How close can symbol codes come to this theoretical bound?
- ↳ How can we construct an optimal code book?

Kraft-McMillan Theorem

(a) $\forall$ B-ary uniquely decodable symbol codes C:

$$r := \sum_{x \in \mathcal{X}} \frac{1}{B^{\ell(x)}} \leq 1 \quad (*)$$ (where $\ell(x) = |C(x)|$)

(b) $\forall \ell: \mathcal{X} \to \mathbb{N}_0$ that satisfy $(*)$: $\exists$ B-ary prefix code C with $|C(x)| = \ell(x)$ $\forall x \in \mathcal{X}$

Proof:

Lemma: $s \in \mathbb{N}_0$, C uniquely dec. sym. code,

$$Y_s := \{\underline{x} \in \mathcal{X}^* \text{ with } |C^*(\underline{x})| = s\}$$

then: $|Y_s| \leq B^s$

(because
- C^* is injective
- $\exists$ B^s bit strings of length s
- if $|Y_s| > B^s$ then $\exists \underline{x}, \underline{x}' \in Y_s$ with $\underline{x} \neq \underline{x}'$ but $C^*(\underline{x}) = C^*(\underline{x}')$

Proof of (a): let $k \in \mathbb{N}$

$$r^k = \left(\sum_{x \in \mathcal{X}} \frac{1}{B^{\ell(x)}}\right)^k = \left(\sum_{x \in \mathcal{X}} B^{-\ell(x)}\right)^k$$

$$= \underbrace{\left(\sum_{x_1 \in \mathcal{X}} B^{-\ell(x_1)}\right)\left(\sum_{x_2 \in \mathcal{X}} B^{-\ell(x_2)}\right) \cdots \left(\sum_{x_k \in \mathcal{X}} B^{-\ell(x_k)}\right)}_{k \text{ factors}}$$

$$r^k = \sum_{x_1 \in \mathcal{X}, x_2 \in \mathcal{X}, \dots, x_k \in \mathcal{X}} B^{-\sum_{i=1}^{k} \ell(x_i)}$$

$$= \sum_{x \in \mathcal{X}^k} B^{-\underbrace{\sum_{i=1}^{k} \ell(x_i)}_{=s}}$$

$$= \sum_{s=0}^{k\ell_{max}} \sum_{x \in Y_s} B^{-s} \quad = \sum_{s=0}^{k\ell_{max}} \underbrace{|Y_s|}_{\leq B^s} B^{-s}$$

$$\leq \sum_{s=0}^{k\ell_{max}} \underbrace{B^s B^{-s}}_{=1} = k\ell_{max} + 1$$

(i) assume (for now): $\mathcal{X}$ is finite

$\Rightarrow \exists \ell_{max}$ s.t. $\ell(x) \leq \ell_{max} \quad \forall x \in \mathcal{X}$

$\Rightarrow r^k \leq k\ell_{max} + 1 \qquad \forall k \in \mathbb{N}$

$\Rightarrow \frac{r^k}{k} \leq \ell_{max} + \frac{1}{k} \xrightarrow{k \to \infty} \ell_{max}$ ($\xrightarrow{k \to \infty} \infty$ if $r > 1$)

$\Rightarrow \boxed{r \leq 1}$

(ii) if $\mathcal{X}$ is countably infinite, e.g., $\mathcal{X} = \mathbb{N}$:

$$r = \sum_{x \in \mathcal{X}} \frac{1}{B^{\ell(x)}} = \sum_{x=1}^{\infty} B^{-\ell(x)} = \lim_{n \to \infty} \underbrace{\sum_{x=1}^{n} B^{-\ell(x)}}_{\leq 1} \leq 1$$

Proof of (b): constructive proof

- sort $\mathcal{X} = \{x, x', x'', \dots\}$ s.t. $\ell(x) \geq \ell(x') \geq \dots$

Algorithm:

- initialize $\xi \leftarrow 1$
- for $x \in \mathcal{X}$ in above order:
  - ↳ update $\xi \leftarrow \xi - B^{-\ell(x)}$ ($\Rightarrow \xi \in [0,1)$)
  - ↳ write $\xi = (0.???)_B$
  - ↳ set $C(x)$ to the $\ell(x)$ bits after "0." (pad with trailing 0s if necessary)

Claim: $C(x)$ is a prefix code

Proof: problem set 2

Example: Simplified Monopoly

x	$\ell(x)$	ξ	$C(x)$
2	3	$1-2^{-3} = (1.000)_2 - (0.001)_2 = (0.111)_2$	111
6	3	$(0.111)_2 - (0.001)_2 = (0.110)_2$	110
3	2	$(0.11)_2 - 2^{-2} = (0.11)_2 - (0.01)_2 = (0.10)_2$	10
4	2	$(0.10)_2 - (0.01)_2 = (0.01)_2$	01
5	2	$(0.01)_2 - (0.01)_2 = 0 = (0.00)_2$	00

$$r = \sum_{x \in \mathcal{X}} B^{-\ell(x)} = 2 \times 2^{-3} + 3 \times 2^{-2} = 4 \times 2^{-2} = 1 \leq 1 \checkmark$$

($=$: $B \leq 2$)

Q: What is the *minimal expected* code word length of a uniquely decodable symbol code?

Minimize: $L := \sum_{x \in \mathcal{X}} \underbrace{p(x)}_{\text{given}} \underbrace{\ell(x)}_{\text{free (but must satisfy K-M)}}$

Strategy:

1) derive lower bound on L
2) show that $\exists$ code that gets close to lower bound

1) Relaxed optimization problem:

minimize $L := \sum_{x \in \mathcal{X}} p(x)\, \ell(x)$

with constraint $\sum_{x \in \mathcal{X}} B^{-\ell(x)} \leq 1$

where $\ell(x) \in \mathbb{R}_{>0}$

Observation: optimal solution will always satisfy $\sum_{x \in \mathcal{X}} B^{-\ell(x)} = 1$

Lagrange multiplier: minimize

$$\mathcal{L} := L + \lambda \sum_{x \in \mathcal{X}} B^{-\ell(x)}$$

over $\ell(x)\ \forall x \in \mathcal{X}$ and over λ

Def: shorthand $q(x) := B^{-\ell(x)}$

$\Rightarrow \ell(x) = -\log_B q(x)$

$$\forall x: \; 0 = \left.\frac{\partial \mathcal{L}}{\partial q(x)}\right|_{q^*} = \left.\frac{\partial}{\partial q(x)}\right|_{q^*} \left[\sum_{x' \in \mathcal{X}} \left(-p(x') \log_B q(x') + \lambda q(x')\right)\right]$$

$$= -p(x) \frac{1}{\ln B \, q^*(x)} + \lambda$$

$$\Rightarrow \quad q^*(x) = \frac{1}{\lambda \ln B} p(x) \propto p(x)$$

Determine λ:
- we want to make $q^*(x)$ as large as possible
- constraint: $\sum_x q^*(x) \leq 1$
- we know: $\sum_x p(x) = 1$

$\Rightarrow$ optimal choice:

$$\boxed{q^*(x) = p(x)}$$

$$\Rightarrow \ell^*(x) = -\log_B p(x)$$

Thus, minimum of the relaxed optimization problem:

$$L^* = \sum_{x \in \mathcal{X}} p(x) \ell^*(x) = \underbrace{-\sum_{x \in \mathcal{X}} p(x) \log_B p(x)}_{=: H_B[p]}$$

"entropy to base B of prob. dist. p"

Result so far: $\forall$ uniquely dec. sym. codes:

$$\boxed{H_B[p] \leq L}$$

$H_B[p]$: entropy; L: expected code word length

Q: How close can we get to $H_B[p]$?

→ Answer: within less than 1 bit per symbol

Proof: return to discrete optimization problem: $\ell(x) \in \mathbb{N} \quad \forall x \in \mathcal{X}$

→ lets set $\ell(x) := \lceil \ell^*(x) \rceil < \ell^*(x) + 1$

⇒ satisfies K-M: $\sum_{x \in \mathcal{X}} B^{-\ell(x)} \leq \sum_{x \in \mathcal{X}} B^{-\ell^*(x)} = 1$

→ then use constructive proof of K-M part (b) to construct a ~~uniquely~~ prefix code C_{Shannon} with $|C_{\text{Shannon}}(x)| = \ell(x) = \lceil \ell^*(x) \rceil < \ell^*(x) + 1$

$$\Rightarrow \boxed{H_B[p] \leq L_{\text{Shannon}} < H_B[p] + 1} \quad (**)$$

Remark: This procedure is called "Shannon Coding"

Remark!: Shannon coding can be suboptimal

↳ Example:

$H_2[p] \approx 2.20$

x	$p(x)$	$\lceil -\log_2 p(x) \rceil$	$C_{\text{Shannon}}(x)$	$\tilde{C}(x)$	$C_{\text{Huffman}}(x)$
2	1/9	$\lceil 3.17 \rceil = 4$	1111	1111	000
6	1/9	4	1110	1110	001
3	2/9	$\lceil 2.17 \rceil = 3$	110	110	10
5	2/9	3	101	10	11
4	1/3	$\lceil 1.58 \rceil = 2$	01	0	01
L			$\frac{26}{9} \approx 2.89$	$\frac{7}{3} \approx 2.33$	$\frac{20}{9} \approx 2.22$

Optimal symbol code: Huffman coding