Information Theory. 3rd Chapter Slides

Boris Kudryashov

ITMO University

December 21, 2016

1. Universal coding task
2. Useful combinatorial formulas
3. Two pass encoding
4. Enumerative coding
5. Asymptotic bounds of redundancy
6. Adaptive coding
7. Algorithm comparison

Universal coding task

- *Encoding redundancy* for a model class Ω is

$$r_n(\Omega) = \sup_{\omega \in \Omega} \left[\bar{R}_n(\omega) - H_\omega \right]. \qquad (1)$$

- Coding is called *Universal* if for algorithm holds

$$\lim_{n \to \infty} r_n(\Omega) = 0,$$

Useful combinatorial formulas

- Consider sequences $\boldsymbol{x} = (x_1, ..., x_n)$, where x_i has one of M_i values, $i = 1, ..., n$. Number of different $\boldsymbol{x}$ is

$$|\{\boldsymbol{x} = (x_1, ..., x_n) : x_i \in \{0, ..., M_i - 1\}, i = 1, ...n\}| = \prod_{i=1}^{n} M_i. \quad (2)$$

-

$$A_M^n = M(M-1) \times ... \times (M-n+1) = \frac{M!}{(M-n)!}. \quad (3)$$

Useful combinatorial formulas

- Number of combinations

$$
\begin{aligned}
C_M^n &= \binom{M}{n} = \frac{A_M^n}{P_n} = \\
&= \frac{M(M-1)\times \ldots \times (M-n+1)}{n!} = \\
&= \frac{M!}{n!(M-n)!}.
\end{aligned} \tag{4}
$$

- Number of combinations

$$
\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!}, & \text{if } n \geq k \geq 0 \\ 1, & \text{if } n \geq 0 \text{ and } k = 0 \text{ or } k = n \\ 0, & \text{if } k < 0 \text{ or } k > n \end{cases} \tag{5}
$$

Useful combinatorial formulas

- binomial coefficient

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

- Number os binary sequences of length n, which contain τ_1 ones and $\tau_0 = n - \tau_1$ zeros.

$$N(\tau_0, \tau_1) = \binom{n}{\tau_0} = \frac{n}{\tau_0! \tau_1!}. \tag{6}$$

- Composition of sequence $\boldsymbol{x}$ is vector $\boldsymbol{\tau}(\boldsymbol{x}) = (\tau_0(\boldsymbol{x}), ..., \tau_{M-1}(\boldsymbol{x}))$, where $\tau_i(\boldsymbol{x})$ denotes number of elements $x_t = i$ in sequence $\boldsymbol{x} = (x_1, ..., x_n)$.

- For arbitrary M

$$N(\boldsymbol{\tau}) = \frac{n!}{\tau_0! \ldots \tau_{M-1}!}. \tag{7}$$

- Newton formula generalization

$$(a_0 + \ldots + a_{M-1})^n = \sum_{\boldsymbol{\tau}:\ \tau_0 + \ldots + \tau_{M-1} = n} N(\boldsymbol{\tau}) \prod_{i=0}^{M-1} a_i^{\tau_i}.$$

- Consider the following lemma:

Lemma

$n \in$ mathbbN can be written as sum of M non-negative integer terms in $\binom{n+M-1}{M-1}$ ways.

- Number of different compositions of sequence of length n over M-size alphabet is

$$N_\tau(n, M) = \binom{n+M-1}{M-1} \tag{8}$$

- Stirling formula

$$\sqrt{2\pi n}n^n e^{-n} \exp\left\{\frac{1}{12n+1}\right\} < n! < \sqrt{2\pi n}n^n e^{-n} \exp\left\{\frac{1}{12n}\right\}. \tag{9}$$

Useful combinatorial formulas

- Consider

$$N(\boldsymbol{\tau}) < (2\pi n)^{-\frac{M-1}{2}} 2^{n\log n - \sum_i \tau_i \log \tau_i} \left(\prod_i \frac{n}{\tau_i} \right)^{1/2} \times$$

$$\times \exp\left\{ \frac{1}{12n} - \sum_i \frac{1}{12\tau_i + 1} \right\}. \tag{10}$$

- Logarithm of number of sequences with specified composition

$$\log N(\boldsymbol{\tau}) < nH(\hat{\boldsymbol{p}}) - \frac{M-1}{2}\log(2\pi n) - \frac{1}{2}\sum_i \log(\hat{p}_i), \tag{11}$$

- More compact estimation

$$\log N(\boldsymbol{\tau}) < nH(\hat{\boldsymbol{p}}) - \frac{M-1}{2}\log(2\pi n) + \frac{1}{2}\log\frac{n}{n-M+1}. \tag{12}$$

- Recurrent formula holds

$$\binom{n+1}{w} = \binom{n}{w} + \binom{n}{w-1}. \tag{13}$$

- $$\binom{n+1}{w} = \binom{n}{w} + \binom{n-1}{w-1} + \ldots + \binom{n-w+1}{1}. \tag{14}$$

- IF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CAN

(15)

- $$l_2 = 6 + 6 + 12 \times 2 + 5 \times 3 + \ldots + 6 = 178.$$

  - 00010000010100110111101101111.

- $$l_1 = 29 + 8 \times 15 = 149 \text{ bit.}$$
- $$l = l_1 + l_2 = 149 + 178 = 327 \text{ bit.} \qquad (16)$$

Table: Huffman code for text (15)

Character	Number of iterations	Codeword length	Codeword
I	1	6	010000
F	1	6	010001
_	12	2	00
W	5	3	100
E	4	4	0101
C	2	5	01001
A	4	4	1010
N	3	4	1011
O	5	3	110
T	1	6	011110
D	4	4	0110
S	3	4	1110
U	2	4	1111
L	2	5	01110
H	1	6	011111

Two pass encoding

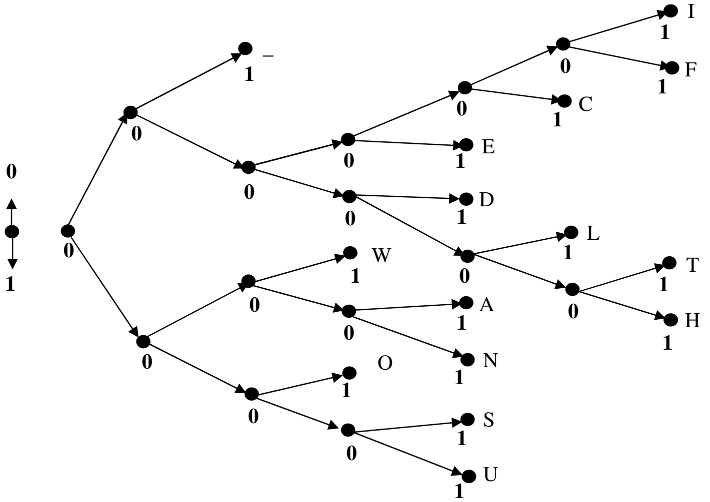

Figure: Huffman codetree for (15)

Table: Regular Huffman code

Character	Codeword length	Codeword
_	2	00
O	3	010
W	3	011
A	4	1000
D	4	1001
E	4	1010
N	4	1011
S	4	1100
U	4	1101
C	5	11100
L	5	11101
F	6	111100
H	6	111101
I	6	111110
T	6	111111

Table: Number of bits for regular code tree transmitting

Level	Number of nodes	Number of leaves n_i	Range of values n_i	Expenses in bits
0	1	0	$0 \dots 1$	1
1	2	0	$0 \dots 2$	2
2	4	1	$0 \dots 4$	3
3	6	2	$0 \dots 6$	3
4	8	6	$0 \dots 8$	4
5	4	2	$0 \dots 4$	3
6	4	4	$0 \dots 4$	3
Всего				19

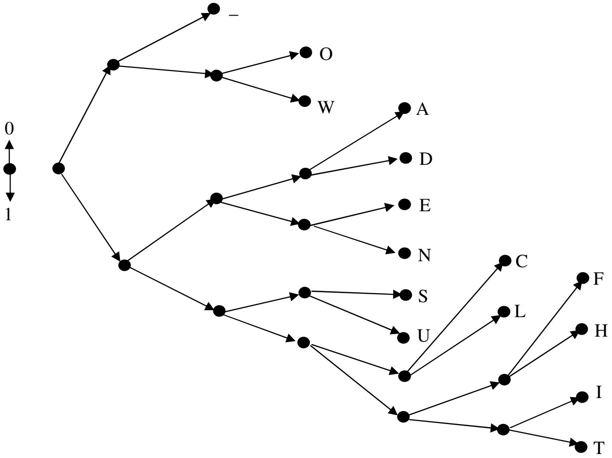

Figure: Codetree for regular Huffman code

Two pass encoding

- Enough for transmitting information about letters that are associated with regular codetree nodes:

$$\left\lceil \log \binom{256}{1} \right\rceil + \left\lceil \log \binom{255}{2} \right\rceil + \left\lceil \log \binom{253}{6} \right\rceil +$$

$$+ \left\lceil \log \binom{247}{2} \right\rceil + \left\lceil \log \binom{245}{4} \right\rceil = 105 \text{ bits}$$

- More precise

$$I = 178 + 19 + 105 = 302 \text{ bits.} \tag{17}$$

Theorem

For two pass coding with Huffman code of Discrete Memoryless Source with alphabet size M and entropy H, average code rate satisfies

$$\bar{R} \le H + 1 + \frac{1}{n}\left(M \log M + 3M - 1\right). \tag{18}$$

Proof.

- $l_1(\boldsymbol{x}) \le 2M - 1 + M \lceil \log M \rceil \le M \log M + 3M - 1.$
-

$$
\begin{aligned}
l_2(\boldsymbol{x}) &\overset{\text{(a)}}{=} \sum_{i=1}^{n} l(x_i) = \\
&\overset{\text{(b)}}{=} \sum_{x \in X} \tau_n(x) l(x) = \\
&\overset{\text{(c)}}{=} n \sum_{x \in X} \frac{\tau_n(x)}{n} l(x) = \\
&\overset{\text{(d)}}{=} n \sum_{x \in X} \hat{p}_n(x) l(x) = \\
&\overset{\text{(e)}}{=} n \mathbf{M}_{\hat{\boldsymbol{p}}_n} \left[l(x) \right] \le \\
&\overset{\text{(f)}}{\le} n(H(\hat{\boldsymbol{p}}_n) + 1).
\end{aligned} \tag{19}
$$

Proof.

- $$\begin{aligned}\bar{R}(\boldsymbol{x}) &= \frac{l(\boldsymbol{x})}{n} = \frac{l_1(\boldsymbol{x}) + l_2(\boldsymbol{x})}{n} \leq & (20)\\ &\leq H(\hat{\boldsymbol{p}}_n) + 1 + \frac{1}{n}\left(M \log M + 3M - 1\right). & (21)\end{aligned}$$

- $$\mathbf{M}\left[H(\hat{\boldsymbol{p}}_n\right] \overset{\text{(a)}}{\leq} H\left(\mathbf{M}\left[\hat{\boldsymbol{p}}_n\right]\right) \overset{\text{(b)}}{=} H(\boldsymbol{p}) = H. \qquad (22)$$

- $$\mathbf{M}\left[\hat{\boldsymbol{p}}_n\right] = \boldsymbol{p}, \qquad (23)$$

- $$\mathbf{M}\left[\frac{\tau_n(a)}{n}\right] = p(a),\ a \in X.$$

Proof.

- $$\chi_a(x) = \begin{cases} 1, & \text{if } x = a, \\ 0, & \text{if } x \neq a. \end{cases}$$

- $$\mathbf{M}\left[\chi_a(x)\right] = 1 \times p(a) + 0 \times (1 - p(a)) = p(a).$$

- $$\begin{aligned} \mathbf{M}\left[\frac{\tau_n(a)}{n}\right] &= \frac{1}{n}\mathbf{M}\left[\sum_{i=1}^{n}\chi_a(x_i)\right] = \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}\left[\chi_a(x_i)\right] = \\ &= p(a), \;\; a \in X. \end{aligned}$$

- Note, that coding redundancy satisfies

$$r = \bar{R} - H \le 1 + \frac{K}{n}, \qquad (24)$$

- When using arithmetic coding, the redundancy can be achieved:

$$r(n) = \frac{M-1}{n} \log n + \frac{K}{n}, \qquad (25)$$

where M alphabet size, K is a constant.

- Codeword length

$$\begin{aligned} l(\boldsymbol{x}) &= l_1(\boldsymbol{x}) + l_2(\boldsymbol{x}) = \\ &= \lceil \log N_\tau(M) \rceil + \lceil \log N(\boldsymbol{\tau}) \rceil = \\ &= \left\lceil \log \binom{n+M-1}{M-1} \right\rceil + \left\lceil \log \frac{n!}{\prod_{i=0}^{M-1} \tau_i(\boldsymbol{x})!} \right\rceil \end{aligned} \qquad (26)$$

- IF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CAN
-

$$\begin{aligned} l_1 &= \left\lceil \log \binom{50+255}{255} \right\rceil = 190 \text{ bits}, \\ l_2 &= \left\lceil \log \left(\frac{50!}{12!5!^2 4!^3 3!^2 2!^3} \right) \right\rceil = 150 \text{ bits}. \end{aligned}$$

- $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_0, ..., \tilde{\tau}_{M-1})$=(12, 5, 5, 4, 4, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 0,. . . ., 0).
- $$1 \le \tilde{\tau}_0 \le n, \quad \tilde{\tau}_{j+1} \le \tilde{\tau}_j, \quad j \ge 0.$$
- It's possible to transmit number of composition, no more than

$$\left\lceil \log \left(n \prod_{j:\tau_j>0} \tau_j \right) \right\rceil \text{ bits} \tag{27}$$

- To transmit the composition $\boldsymbol{\tau}' = (1, 2, 3, 2, 3, 4, 241)$

$$l_1 = \left\lceil \log \left(n \prod_{j:\tau_j>0} \tau_j \right) \right\rceil + \left\lceil \log \left(\frac{M!}{\prod_j \tau_j'!} \right) \right\rceil = 25 + 108 = 133 \text{ bits.} \tag{28}$$

- $$l = l_1 + l_2 = 283 \text{ bits}$$

- $$\begin{aligned} l(\boldsymbol{x}) &= l_1(\boldsymbol{x}) + l_2(\boldsymbol{x}) = \\ &= \lceil \log N_\tau(M) \rceil + \lceil \log N(\boldsymbol{\tau}) \rceil \le \\ &\le nH(\hat{\mathbf{p}}) - \frac{M-1}{2}\log(2\pi n) - \\ &\quad -\frac{1}{2}\sum_i \log(\hat{p}_i) + (M-1)\log(n+1) + 1. \end{aligned}$$

Theorem

For enumerate coding of discrete memoryless source with alphabet size M and entropy H, the average code rate is.

$$\bar{R} \leq H + \frac{M-1}{2}\frac{\log(n+1)+K}{n}, \qquad (29)$$

where K does not depend on sequence length n.

Example

- Let $n = 10$, $\boldsymbol{\tau} = (2, 5, 3)$, $\boldsymbol{x} = (2011021211)$. Probability distribution on first step $\boldsymbol{\tau}/n = (2/10, 5/10, 3/10)$ $G = 0.3$.
- Probability distribution after first step: $(2, 5, 2)/9 = (2/9, 5/9, 2/9)$.
- After second step: $G = 3/10 \times 2/9$.
- After 10 steps:

$$G = \frac{3}{10} \times \frac{2}{9} \times \frac{5}{8} \times \frac{4}{7} \times \frac{1}{6} \times \frac{2}{5} \times \frac{3}{4} \times \frac{1}{3} \times \frac{2}{2} \times \frac{1}{1} \quad .$$

- Codeword length for this example:

$$l = \lceil -\log G \rceil = \left\lceil -\log \frac{10!}{2!5!3!} \right\rceil$$

Asymptotic bounds of redundancy

- Redundancy

$$r_n(\Omega) = \sup_{\omega \in \Omega} \left[\bar{R}_n(\omega) - H_\omega \right] . \qquad (30)$$

- For a given $\boldsymbol{\theta}$

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=0}^{M-1} \theta_i^{\tau_i(\boldsymbol{x})},$$

where $\boldsymbol{\tau}(\boldsymbol{x}) = (\tau_0(\boldsymbol{x}), \ldots, \tau_{M-1}(\boldsymbol{x}))$ composition of sequence $\boldsymbol{x}$.

Asymptotic bounds of redundancy

- $$r_n(\Theta) \leq \frac{M-1}{2}\frac{\log(n+1)+K}{n}, \qquad (31)$$
  where K does not depend on n.
- $$r_n(\Theta) = \inf \sup_{\theta \in \Theta} \left[\bar{R}_n(\boldsymbol{\theta}) - H(X|\boldsymbol{\theta}) \right], \qquad (32)$$
  where $H(X|\boldsymbol{\theta})$ is entropy of X.

Asymptotic bounds of redundancy

Theorem

For a discrete memoryless source with alphabet size M, with redundancy of universal code $r_n(\Theta)$ per message of length n holds:

$$r_n(\Theta) \geq \frac{M-1}{2} \frac{\log n + C}{n}, \tag{33}$$

where C does not depend on n.

Proof of theorem.

- Consider $\Theta = \{\boldsymbol{\theta}\}$, $f(\boldsymbol{\theta})$. Required redundancy:

$$r_n(\Theta) = \inf_{f(\boldsymbol{\theta})} \sup \mathbf{M}_f \left[\bar{R}_n(\boldsymbol{\theta}) - H(X|\boldsymbol{\theta})\right], \quad (34)$$

-

$$r_n(\Theta) \geq \sup_{f(\boldsymbol{\theta})} \inf \mathbf{M}_f \left[\bar{R}_n(\boldsymbol{\theta}) - H(X|\boldsymbol{\theta})\right], \quad (35)$$

-

$$\sum_{\boldsymbol{x} \in X^n} 2^{-l(\boldsymbol{x})} \leq 1.$$

Proof of theorem

- $$q(\boldsymbol{x}) = 2^{-l(\boldsymbol{x})},$$

- $$-\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}|\boldsymbol{\theta}) \log q(\boldsymbol{x}) \le \bar{l}_n(q,\boldsymbol{\theta}) \le -\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}|\boldsymbol{\theta}) \log q(\boldsymbol{x}) + 1.$$

- $$\bar{l}_n(q,\boldsymbol{\theta}) = -\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}|\boldsymbol{\theta}) \log q(\boldsymbol{x}). \qquad (36)$$

Proof of theorem

- $$\bar{I}_n(q) = -\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}) \log q(\boldsymbol{x}), \tag{37}$$

- $$p(\boldsymbol{x}) = \mathbf{M}_f\left[p(\boldsymbol{x}|\boldsymbol{\theta})\right] = \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) p(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{38}$$

Proof of theorem

- Minimum at right is achieved, when $p(\boldsymbol{x}) = q(\boldsymbol{x})$.

$$-\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}) \log p(\boldsymbol{x}) + \sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}) \log q(\boldsymbol{x}) = -L(p||q) \le 0$$

-

$$r_n(\Theta) \ge \frac{1}{n} \sup_{f(\boldsymbol{\theta})} \mathbf{M}_f \sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}|\boldsymbol{\theta}) \log \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})} \quad . \qquad (39)$$

Proof of theorem

- Dirichlet distribution:

$$f_{\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \Gamma\left(\sum_{i=0}^{M-1} \lambda_i\right) \prod_{i=0}^{M-1} \frac{\theta_i^{\lambda_i - 1}}{\Gamma(\lambda_i)}, \tag{40}$$

  where $\boldsymbol{\lambda} = (\lambda_0, ..., \lambda_{M-1})$ is vector of distribution parameters, $\lambda_i \geq 0$, $i = 0, ..., M-1$, $\Gamma(z)$ is Gamma function.

- Gamma function:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt.$$

- $$\Gamma(x) = (x-1)\Gamma(x-1), \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Proof of theorem

- $$\Gamma(n) = (n-1)! \quad .$$
- $$\Gamma(z) \approx \sqrt{2\pi} z^{z-1/2} e^{-z}, \quad z \to \infty.$$
- For some K_1

$$\left|\log \Gamma(z) + z \log e - (z - 1/2) \log z\right| \leq K_1. \tag{41}$$

Asymptotic bounds of redundancy

Proof of theorem

- Dirichlet integral: $\forall \alpha_i$, $i = 1, ..., n$ and a continuous function f:

$$\int_{\boldsymbol{x}:\sum_i x_i=1} f\left(\sum_{i=1}^{n} x_i\right) \prod_{i=1}^{n} x_i^{\alpha_i-1} dx_1 ... dx_n =$$

$$= \frac{\prod_{i=1}^{n} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{n} \alpha_i)} \int_0^1 f(\tau) \tau^{\left(\sum_{i=1}^{n} \alpha_i\right)-1} d\tau. \quad (42)$$

- Consider parameter $\lambda_i = 1/2$, $i = 1, ..., M$,

$$f(\boldsymbol{\theta}) = \frac{\Gamma(M/2)}{\pi^{M/2}} \prod_{i=0}^{M-1} \theta_i^{-1/2}. \quad (43)$$

Asymptotic bounds of redundancy

Proof of theorem

- $$p(\mathbf{x}) = \frac{\Gamma(M/2)}{\pi^{M/2}} \int_{\boldsymbol{\theta}} \prod_{i=0}^{M-1} \theta_i^{\tau_i(\mathbf{x})-1/2} d\boldsymbol{\theta}.$$
- $$p(\mathbf{x}) = \frac{\Gamma(M/2)}{\pi^{M/2}} \quad \frac{\prod_{i=0}^{M-1} \Gamma(\tau_i(\mathbf{x}) + 1/2)}{\Gamma(n + M/2)}. \quad (44)$$
- $$-\log p(\mathbf{x}) = nH\left(\frac{\boldsymbol{\tau}(\mathbf{x})}{n}\right) + \frac{M-1}{2} \log n + K(n), \quad (45)$$

  where $K(n)$ is bounded

Lemma

For a given parameters θ, the average "empirical entropy" $H\left(\frac{\tau(x)}{n}\right)$ $\forall\theta$ is connected to it's entropy $H_{\theta}(X)$ by inequalities:

$$-\frac{K_1}{n} \leq \sum_{x} p(x|\theta) H\left(\frac{\tau(x)}{n}\right) - H_{\theta}(X) \leq 0 \quad , \tag{46}$$

where K_1 is a constant, $H_{\theta}(X) = -\sum_{i=0}^{M-1} \theta_i \log \theta_i$ is entropy of the source

Proof of Lemma

- $$\sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{\theta})\tau_i(\boldsymbol{x}) = n\theta_i \quad , \qquad (47)$$

- $$\sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{\theta})\tau_i^2(\boldsymbol{x}) = n^2\theta_i^2 + n\theta_i(1-\theta_i) \quad . \quad (48)$$

Proof of Lemma

$$
\begin{aligned}
& -\sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{\theta}) \sum_{i=0}^{M-1} \frac{\tau_i(\boldsymbol{x})}{n} \log \frac{\tau_i(\boldsymbol{x})}{n} + \sum_{i=0}^{M-1} \theta_i \log \theta_i = \\
\stackrel{\text{(a)}}{=} \quad & -\sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{\theta}) \sum_{i=0}^{M-1} \frac{\tau_i(\boldsymbol{x})}{n} \log \frac{\tau_i(\boldsymbol{x})}{n} + \sum_{\boldsymbol{x}} \sum_{i=0}^{M-1} p(\boldsymbol{x}|\boldsymbol{\theta}) \frac{\tau_i(\boldsymbol{x})}{n} \log \theta_i = \\
= \quad & -\sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{\theta}) \sum_{i=0}^{M-1} \frac{\tau_i(\boldsymbol{x})}{n} \log \frac{\tau_i(\boldsymbol{x})}{n\theta_i} \geq \\
\stackrel{\text{(b)}}{\geq} \quad & -\log \sum_{\boldsymbol{x}} \sum_{i=0}^{M-1} p(\boldsymbol{x}|\boldsymbol{\theta}) \frac{\tau_i^2(\boldsymbol{x})}{n^2\theta_i} = \\
\stackrel{\text{(c)}}{=} \quad & = -\log \sum_{i=0}^{M-1} \frac{n^2\theta_i^2 + \theta_i(1-\theta_i)}{n^2\theta_i} = \\
= \quad & -\log\left(1 + \frac{M-1}{n}\right) \geq \frac{M-1}{n} \log e.
\end{aligned}
$$

Proof of theorem

- $$-\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}|\boldsymbol{\theta})\log p(\boldsymbol{x}|\boldsymbol{\theta}) = H(X^n|\boldsymbol{\theta}) = nH_{\boldsymbol{\theta}}(X).$$

- $$\sum_{\boldsymbol{x}\in X^n} p(\boldsymbol{x}|\boldsymbol{\theta})\log\frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})} \geq \frac{M-1}{2}\log n + K(n) - \frac{K_1}{n}.$$

Proof of theorem

- $$\begin{aligned} -\log p(\boldsymbol{x}) &= -\sum_{i=0}^{M-1} \tau_i(\boldsymbol{x}) \log\left(\tau_i(\boldsymbol{x}) + \frac{1}{2}\right) + \\ &\quad + \left(n + \frac{M-1}{2}\right) \log\left(n + \frac{M}{2}\right) + K_2 = \\ &= -\sum_{i=0}^{M-1} \tau_i(\boldsymbol{x}) \log\left[\frac{\tau_i(\boldsymbol{x})}{n}\left(1 + \frac{1}{2\tau_i(\boldsymbol{x})}\right)\right] - n\log n + \\ &\quad + \left(n + \frac{M-1}{2}\right) \log n + \\ &\quad + \left(n + \frac{M-1}{2}\right) \log\left(1 + \frac{M}{2n}\right) + K_2, \end{aligned}$$
- $$0 \le \log(1+\epsilon) \le \epsilon \log e,$$

Example 1

- IF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CAN
- $$G = 1 \cdot \frac{1}{256} \cdot \frac{1}{257} \cdot \frac{1}{258} \cdot \frac{1}{259} \cdot \frac{1}{260} \cdot \frac{2}{261} \cdot \ldots =$$
$$= \frac{12!(5!)^2(4!)^3(3!)^2(2!)^3}{256 \cdot \ldots \cdot 305}.$$
- $$I = \lceil -\log G \rceil + 1 = 342 \text{ bits.} \tag{49}$$
- $$\hat{p}_n(a) = \frac{\tau_n(a) + 1/2}{n + M/2} = \frac{2\tau_n(a) + 1}{2n + M}. \tag{50}$$

Example 2

- $$G = 1 \cdot \frac{1}{256} \cdot \frac{1}{258} \cdot \frac{1}{260} \cdot \frac{1}{262} \cdot \frac{1}{264} \cdot \frac{3}{266} \cdot \ldots =$$
$$= \frac{23!!(9!!)^2(7!!)^3(5!!)^2(3!!)^3}{256 \cdot \ldots \cdot 305}.$$
- $$(2n-1)!! = 1 \times 3 \times ... \times (2n-1),$$
$$(2n)!! = 2 \times 4 \times ... \times (2n).$$
- $$l = \lceil -\log G \rceil + 1 = 323 \text{ bits.} \qquad (51)$$

- $$\hat{p}_n(a) = \frac{\tau_n(a)}{n+1}, \quad \tau_n(a) > 0.$$
- $$\hat{p}_n(esc) = \frac{1}{n+1},$$
- *A-algorithm* formula

$$\hat{p}_n(a) = \begin{cases} \frac{\tau_n(a)}{n+1}, & \text{if } \tau_n(a) > 0; \\ \frac{1}{n+1}\frac{1}{M-M_n}, & \text{if } \tau_n(a) = 0, \end{cases} \tag{52}$$

- IF_WE_CANNOT_DO_AS_WE_WOULD_WE_SHOULD_DO_AS_WE_CAN
- $$G = 1 \cdot \frac{1}{2} \cdot \frac{1}{256} \cdot \frac{1}{3} \cdot \frac{1}{255} \cdots$$
- $$G = \frac{11!(4!)^2(3!)^3(2!)^2}{50!} \cdot \frac{1}{256 \cdot 255 \cdot \ldots \cdot 242}.$$

$$I = \lceil -\log G \rceil + 1 = 291 \text{ bits.} \tag{53}$$

- for a $\boldsymbol{x} = (x_1, ..., x_n)$ with composition $(\tau_1, ..., \tau_M)$

$$
\begin{aligned}
G &= \frac{\prod_{i=1}^{M_n} (\tau_i - 1)!}{n!} \cdot \frac{(M - M_n)!}{M!} = \\
&= \frac{\prod_{i=1}^{M_n} \tau_i!}{n!} \cdot \frac{(M - M_n)!}{M! \prod_{i=1}^{M_n} \tau_i} \geq \\
&\geq \frac{\prod_{i=1}^{M_n} \tau_i!}{n!} M^{-M} (n+1)^{-M}.
\end{aligned}
$$

Theorem

For adaptive arithmetic coding of discrete memoryless source with alphabet size size M and entropy H, average code rate satisfies

$$\bar{R} \leq H + \frac{M}{2} \frac{\log(n+1) + K}{n}, \tag{54}$$

where K does not depend on n.

Example 4. D-algorithm

- $$\hat{p}_n(a) = \begin{cases} \frac{\tau_n(a)-1/2}{n}, & \text{if } \tau_n(a) > 0; \\ \frac{M_n}{2n}\frac{1}{M-M_n}, & \text{if } \tau_n(a) = 0. \end{cases} \quad (55)$$

- $$G = 1 \cdot \frac{1}{2} \cdot \frac{1}{256} \cdot \frac{2}{4} \cdot \frac{1}{255} \cdot \frac{3}{6} \cdot \frac{1}{254} \frac{4}{8} \cdot \frac{1}{253} \frac{1}{10} \cdot \frac{5}{12} \cdot \frac{1}{252} \ldots$$

Example 4. D-algorithm

- $$(2 \times 4 \times \ldots \times 100) \times (256 \times 255 \times \ldots \times 242).$$

- $$G = \frac{(2 \times 12 - 3)!!((2 \times 5 - 3)!!)^2((2 \times 4 - 3)!!)^3((2 \times 3 - 3)!!)^2}{100!!} \times$$
$$\times \frac{14!}{256 \cdot 255 \cdot \ldots \cdot 242}.$$

- $$I = \lceil -\log G \rceil + 1 = 283 \quad \text{bits}.$$

Table: Universal coding algorithm comparison

Algorithm	Number of traverses	Asymptotic redundancy	codeword length for text (15)
2-traverse coding, Huffman code	2	$1 + K_1/n$	302
Enumerative coding	2	$\frac{M \log n + K_3}{2n}$	283
Adaptive coding (A)	1	$\frac{M \log n + K_4}{2n}$	291
Adaptive coding (D)	1	$\frac{M \log n + K_5}{2n}$	283