UNIVERSITEIT VAN AMSTERDAM

INFORMATIE EN COMMUNICATIE

HW-2

Authors:
Abe WIERSMA

January 13, 2016

1. Let X be a discrete random variable. Show that the entropy of a function g of X is less than or equal to the entropy of X by justifying the following steps:

$H[g(X)|X] = 0$ the value of g(X) is completely determined by the value of X.

$$H(X) = H(X) + H[g(X)|X] \tag{1}$$

$H(g(X)|X) = H(X, g(X)) - H(X)$ Shannon Entropy conditional chain rule

$$= H(X, g(X)) \tag{2}$$

H(X,g(X)) = H(g(X),X), apply Shannon Entropy conditional chain rule again, $H(g(X), X) = H(X|g(X)) + H(g(X))$

$$= H(g(X)) + H(X|g(X)) \tag{3}$$

$H(g(X)) + H(X|g(X)) \geq H(g(X))$ **Proposition 4[CF]:** $\implies H(X|g(X)) \geq 0$

$$\geq H(g(X)) \tag{4}$$

2. Let X and Y be independent binary random variables with:

$$P_X[1] = P_X[0] = P_Y[1] = P_Y[0] = \frac{1}{2}$$

Compute H(X+Y).

$$Z = X + Y \tag{5}$$

$$p(Z) = \begin{cases} \frac{1}{4} & Z = 0 \\ \frac{1}{2} & Z = 1 \\ \frac{1}{4} & Z = 2 \end{cases} \tag{6}$$

$$H(Z) = -\left(\sum_{z \in Z} P(z) \log P(z)\right) \tag{7}$$

$$= -\left(\frac{1}{4}log\left(\frac{1}{4}\right) + \frac{1}{2}log\left(\frac{1}{2}\right) + \frac{1}{4}log\left(\frac{1}{4}\right)\right) \tag{8}$$

$$= 1.5 \tag{9}$$

3. Use Jensen's inequality to derive an inequality between $E[X^2]$ and $E[X]^2$. Use this inequality as an alternative proof that $Var[X] \geq 0$.

Proposition 2. Jensen's inequality

$$E\left[f(X)\right] \geq f\left(E[X]\right) \tag{10}$$

Define $f(x)$

$$f(x) = x^2 \tag{11}$$

Apply Jensen's inequality

$$E\left[X^2\right] \geq \left(E[X]\right)^2 \tag{12}$$

Show $Var[X] \geq 0$

$$\begin{align} E\left[X^2\right] - \left(E[X]\right)^2 &\geq 0 \tag{13} \\ Var[X] &\geq 0 \tag{14} \end{align}$$

4. The mutual information between two random variables X and Y is defined as $I(X;Y) := H(X) - H(X|Y)$

   (a) Show that the mutual information can be expressed in terms of the relative entropy, i.e. that $I(X;Y) = D_{KL}(P_{XY}||P_X P_Y)$

$$\begin{align} I(X;Y) &= H(X) - H(X|Y) \tag{15} \\ &= -\sum_{x\in X} P_X(x)\log P_X(x) - \sum_{x\in X, y\in Y} P_{XY}(x,y)\log\frac{P_Y(y)}{P_{XY}(x,y)} \tag{16} \\ &= -\sum_{x\in X}\sum_{y\in Y} P_X(x)P_{XY}(y|x)\log P_X(x) - \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)\log\frac{P_Y(y)}{P_{XY}(x,y)} \tag{17} \\ &= -\sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)\log\frac{P_Y(y)}{P_{XY}(x,y)} - \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)\log P_X(x) \tag{18} \\ &= \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)\left(-\log\frac{P_Y(y)}{P_{XY}(x,y)} - \log P_X(x))\right) \tag{19} \\ &= \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)(\log P_{XY}(x,y) - (\log P_X(x) + \log P_Y(y))) \tag{20} \\ &= \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)(\log P_{XY}(x,y) - \log P_X(x)P_Y(y)) \tag{21} \\ &= \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)\log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \tag{22} \\ &= D_{\mathrm{KL}}(P_{XY}\|P_X P_Y) \tag{23} \end{align}$$

(b) Use (a) and Class exercise 6 to prove that $H(X|Y) \leq H(X)$

$$H(X|Y) \leq H(X) \tag{24}$$
$$H(X) - H(X|Y) \geq 0 \tag{25}$$

use (a)

$$D_{\mathrm{KL}}(P_{XY} \| P_X P_Y) \geq 0 \tag{26}$$

use Class exercise 6

$$P_X = P_{XY}, Q_X = P_X P_Y \tag{27}$$

✓

5. **Kraft's Inequality:** Below, six binary codes are shown for the source symbols $x_1 \dots x_4$.

	Code A	Code B	Code C	Code D	Code E	Code F
x_1	00	0	0	0	1	1
x_2	01	10	11	100	01	10
x_3	10	11	100	110	001	100
x_4	11	110	110	111	0001	1000

Kraft's inequality There exists a prefix-free code $(C) = c_1, \dots, c_m)$ with codeword lenghts $l(c_1) = l_1, \dots, l(c_m) = l_m \in \mathbb{N}_0$ if and only if

$$\sum_{i=1}^{m} \frac{1}{2^{l_i}} \leq 1 \tag{28}$$

(29)

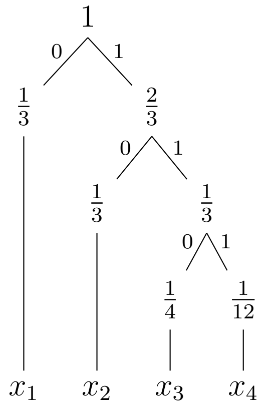
(a) Which codes fulfill the Kraft inequality?
Code A $= 4(\frac{1}{4}) = 1 \leq 1$ ✓
Code B $= \frac{1}{2} + 2(\frac{1}{4}) + \frac{1}{8} = \frac{9}{8} \leq 1$✗
Code C $= \frac{1}{2} + \frac{1}{4} + 2(\frac{1}{8}) = 1 \leq 1$✓
Code D $= \frac{1}{2} + 3(\frac{1}{8}) = \frac{7}{8} \leq 1$✓
Code E $= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16} \leq 1$✓
Code F $= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16} \leq 1$✓

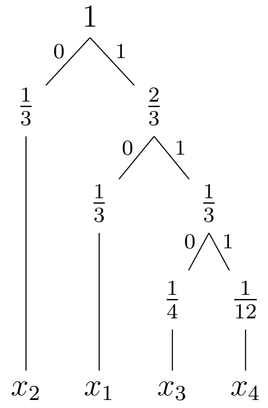
(b) Is a code that satisfies this inequality always uniquely decodable?
No, ex: Code C, $x_4 = x_2 + x_1$

(c) Which codes are prefix-free codes?
A, D, E
!B x_3 is prefix x_4, !C x_2 is prefix x_4, !F x_1 is prefix x_2, x_3, x_4

(d) Which codes are uniquely decodable?
Prefix-free $\implies$ uniquely decodable.
So A, D, E are already implied uniquely decodable.
$x_4 = x_3 + x_1 \implies !B$, $x_4 = x_2 + x_1 \implies !C$
$E = F^{-1} \implies$ uniquely decodable

6. **Optimal Huffman coding:** Consider a random variable X that takes on four values with probabilities $\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12}$. Show that there exist two different sets of optimal length for the (binary) Huffman codewords.





X	C
x_1	0
x_2	10
x_3	110
x_4	111

X	C
x_1	10
x_2	0
x_3	110
x_4	111

Constructed two huffman codes for X, two possibilities

7. **Huffman Coding:** Jane, a student, regularly sends a message to her parents via a binary channel. The binary channel is lossless (i.e. error-free), but the per-bit costs are quite high, so she wants to send as few bits as possible. Each time, she selects one message out of a finite set of possible messages and sends it over the channel. There are 7 possible messages:

   (a) "Everything is fine"
   (b) "I am short on money; please send me some"
   (c) "I'll come home this weekend"
   (d) "I am ill, please come and pick me up"
   (e) "My study is going well, I passed an exam (. . . and send me more money)"
   (f) "I have a new boyfriend"
   (g) "I have bought new shoes"

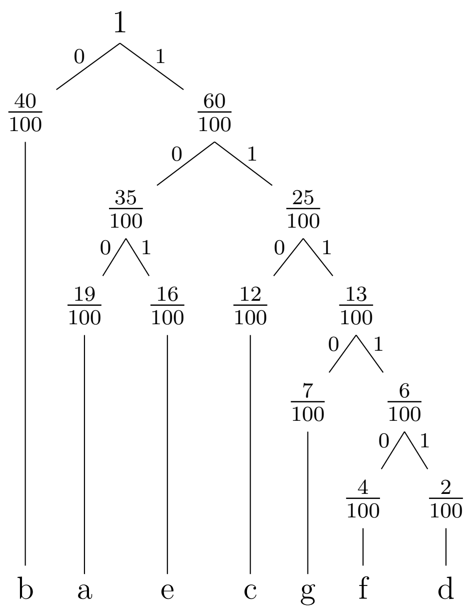
   Based on counting the types of 100 of her past messages, the empirical probabilities of the different messages are:

m	a	b	c	d	e	f	g
$P_M(m)$	$\frac{19}{100}$	$\frac{40}{100}$	$\frac{12}{100}$	$\frac{2}{100}$	$\frac{16}{100}$	$\frac{4}{100}$	$\frac{7}{100}$

   Jane wants to minimize the average number of bits needed to communicate to her parents (with respect to the empirical probability model above).

   (a) Design a Huffman code for Jane and draw the binary tree that belongs to it.

X	C
x_1	0
x_2	10
x_3	110
x_4	111



(b) For a binary source X with $PX(0) = \frac{1}{8}$ and $PX(1) = \frac{7}{8}$, design a Huffman code for blocks of N=1,2 and 3 bits. For each of the three codes, compute the average codeword length and divide it by N, in order to compare it to the optimal length, i.e. the entropy of the source. What do you observe?

N=1	P(x)	C
1	$\frac{7}{8}$	0
0	$\frac{1}{8}$	1

$l_{N_1} = \frac{1}{8} + \frac{7}{8} = 1$
$1/1 = 1$
$H(X) =$
$-(\frac{1}{8}*log(\frac{1}{8})+\frac{7}{8}*log(\frac{7}{8})) \approx 0.5$

N=2	P(x)	C
11	$\frac{49}{64}$	0
01	$\frac{7}{64}$	10
10	$\frac{7}{64}$	110
00	$\frac{1}{64}$	111

$l_{N_2} = \frac{49}{64}+2(\frac{7}{64})+3(\frac{7}{64}+\frac{1}{64}) = \frac{87}{64} \approx 1.4$
$\frac{87}{64}/2 = \frac{87}{128} \approx 0.7$
$H(X) =$
$-((\frac{49}{64} * log(\frac{49}{64})) + 2(\frac{7}{64} * log(\frac{7}{64}))+(\frac{1}{64}*log(\frac{1}{64}))) \approx 1.1$

N=3	P(x)	C
111	$\frac{343}{512}$	0
110	$\frac{14}{512}$	110
011	$\frac{14}{512}$	100
101	$\frac{14}{512}$	101
100	$\frac{7}{512}$	11100
010	$\frac{7}{512}$	11101
001	$\frac{7}{512}$	11110
000	$\frac{1}{512}$	11111

$l_{N_3} = \frac{343}{512}+3(\frac{14}{512}+\frac{14}{512}+\frac{14}{512})+ 5(\frac{7}{512}+\frac{7}{512}+\frac{7}{512}+\frac{1}{512}) = \frac{579}{512} \approx 1.1$
$\frac{579}{512}/3 = \frac{193}{512} \approx 0.4$
$H(X) = -((\frac{343}{512} * log(\frac{343}{512})) + 3(\frac{14}{512} * log(\frac{14}{512})) + 3(\frac{7}{512} * log(\frac{7}{512})) + (\frac{1}{512} * log(\frac{1}{512}))) \approx 1.1$

The code length/N is reducing with 0.3 every +1 increase.

(c) If you were asked at (b) to design a Huffman code for a block of $N = 100$ bits, what problem would you run into?

The code length would become really big for every $X \neq 1^n$