UNIVERSITÉ DE GENÈVE

ANALYSE ET TRAITEMENT DE L'INFORMATION
14X026

TP 7: Entropy and Detection Theory

Author: Lea Heiniger

E-mail: Lea.Heiniger@unige.ch

December 2023

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'informatique

Exercise 1. Quantifiers of information

U	V	W	$p_{U,V,W}(u,v,w)$
0	0	0	$\frac{1}{4}$
0	0	1	0
0	1	0	$\frac{1}{4}$
0	1	1	$\frac{1}{8}$
1	0	0	0
1	0	1	$\frac{1}{8}$
1	1	0	0
1	1	1	$\frac{1}{4}$

Table 1: Caption

From the joint probability mass function we can deduce :

$p(U=0) = \frac{5}{8}$ and $p(U=1) = \frac{3}{8}$
$p(V=0) = \frac{3}{8}$ and $p(V=1) = \frac{5}{8}$
$p(W=0) = \frac{1}{2}$ and $p(W=1) = \frac{1}{2}$

$p(U=0|V=0) = \frac{2}{3}$, $p(U=0|V=1) = \frac{3}{5}$, $p(U=1|V=0) = \frac{1}{3}$ and $p(U=1|V=1) = \frac{2}{5}$
$p(V=0|U=0) = \frac{2}{5}$, $p(V=0|U=1) = \frac{1}{3}$, $p(V=1|U=0) = \frac{3}{5}$ and $p(V=1|U=1) = \frac{2}{3}$
$p(W=0|U=0) = \frac{7}{8}$, $p(W=0|U=1) = 0$, $p(W=1|U=0) = \frac{1}{3}$ and $p(W=1|U=1) = \frac{3}{5}$

1. Calculate $H(U)$, $H(V)$ and $H(W)$
   $$H(U) = -\sum_{u\in\{0,1\}} p_U(u)\log_2(p_U(u)) = 0.9544$$
   $$H(V) = -\sum_{v\in\{0,1\}} p_V(v)\log_2(p_V(v)) = 0.9544$$
   $$H(W) = -\sum_{w\in\{0,1\}} p_W(w)\log_2(p_W(w)) = 1$$

2. Calculate $H(U|V)$, $H(V|U)$ and $H(W|U)$
   $$H(U|V) = -\sum_{u\in\{0,1\},v\in\{0,1\}} p_{U,V}(u,v)\log_2(p_{U|V}(u|v)) = 0.9512$$
   $$H(V|U) = -\sum_{v\in\{0,1\},u\in\{0,1\}} p_{V,U}(v,u)\log_2(p_{V|U}(v|u)) = 0.9512$$
   $$H(W|U) = -\sum_{w\in\{0,1\},u\in\{0,1\}} p_{W,U}(w,u)\log_2(p_{W|U}(w|u)) = 0.5708$$

3. Calculate $I(U;V)$, $I(U;W)$ and $I(U;V;W)$

   $I(U;V) = H(U) - H(U|V) = 0.0032$
   $I(U;W) = H(W) - H(W|U) = 0.4292$
   $$I(U;V;W) = \sum_{u\in\{0,1\},v\in\{0,1\},w\in\{0,1\}} p_{U,V,W}(u,v,w)\log_2\left(\frac{p_{U,V,W}(u,v,w)}{p_U(u)p_V(v)p_W(w)}\right) = 0.6589$$

4. Calculate $H(U,V,W)$
   We can compute $H(W|U,V) = 0.3444$
   $H(U,V,W) = H(U) + H(V|U) + H(W|U,V) = 2.25$

Exercise 2. Source coding

For the implementation see file `tp7.py`

Here are a few outputs I got on various sequences :

- The length of the encoded sequence is : 2004
  if we divide it by the length of the sequence we have : 1.002
  The entropy of the sequence is : 1.0450616314422685

- The length of the encoded sequence is : 2016
  if we divide it by the length of the sequence we have : 1.008
  The entropy of the sequence is : 1.0567216953965506

- The length of the encoded sequence is : 1999
  if we divide it by the length of the sequence we have : 0.9995
  The entropy of the sequence is : 1.039529608622848

We can see that the entropy of the sequence and the number of symbols in the encoded version relative to the initial number of symbols are very close. To make sure I tried to change the probability p of getting a 1 in the sequence. The results (below) confirmed the hypotesis that they are linked.

$p = 0.2$: The length of the encoded sequence is : 2246
if we divide it by the length of the sequence we have : 1.123
The entropy of the sequence is : 1.2487849530919453

$p = 0.4$: The length of the encoded sequence is : 2492
if we divide it by the length of the sequence we have : 1.246
The entropy of the sequence is : 1.4557170900403817

$p = 0.7$: The length of the encoded sequence is : 2372
if we divide it by the length of the sequence we have : 1.186
The entropy of the sequence is : 1.3819702062967405