

COMP2610 / COMP6261 - Information Theory

Tutorial 2: Entropy and Information

Robert Williamson

Semester 2, 2018

1. Let X be a random variable with possible outcomes $\{1, 2, 3\}$. Let the probabilities of the outcomes be

$$\begin{aligned}p(X = 1) &= \frac{\theta}{2} \\p(X = 2) &= \frac{\theta}{2} \\p(X = 3) &= 1 - \theta\end{aligned}$$

for some parameter $\theta \in [0, 1]$.

Suppose we see N observations of the random variable, $\{x_1, \dots, x_N\}$. Let n_i denote the number of times that we observe the outcome $X = i$, i.e.

$$n_i = \sum_{k=1}^N 1_{\{x_k = i\}}$$

- (a) Write down the log-likelihood function in terms of n_1, n_2, n_3 .
(b) Suppose the observations are $\{3, 3, 1, 2, 3\}$.

Compute the maximum likelihood estimate of θ . (Check when the derivative is zero.)

2. Consider the following joint distribution over X, Y :

$p(X, Y)$		X			
		1	2	3	4
Y	1	0	0	1/8	1/8
	2	1/8	1/16	1/16	0
	3	1/8	1/8	0	0
	4	0	1/16	1/16	1/8

- (a) Show that X and Y are not statistically independent. (*Hint*: You need only show that for at least one specific x, y pair, $p(X = x, Y = y) \neq p(X = x)p(Y = y)$.)
(b) Compute the following quantities:
(i) $H(X)$
(ii) $H(Y)$
(iii) $H(X|Y)$
(iv) $H(Y|X)$
(v) $H(X, Y)$
(vi) $I(X; Y)$.

3. A standard deck of cards contains 4 *suits* — $\heartsuit, \diamondsuit, \clubsuit, \spadesuit$ (“hearts”, “diamonds”, “clubs”, “spades”) — each with 13 *values* — A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (The *A, J, Q, K* are called “Ace”, “Jack”, “Queen”, “King”). Each card has a *colour*: hearts and diamonds are coloured red; clubs and spades are black. Cards with values J, Q, K are called *face cards*.

Each of the 52 cards in a deck is identified by its value v and suit s and denoted vs . For example, $2\heartsuit$, $J\clubsuit$, and $7\spadesuit$ are the “two of hearts”, “Jack of clubs”, and “7 of spades”, respectively. The variable c will be used to denote a card’s colour. Let $f = 1$ if a card is a face card and $f = 0$ otherwise.

A card is drawn at random from a thoroughly shuffled deck. Calculate:

- The information $h(c = \text{red}, v = K)$ in observing a red King
 - The conditional information $h(v = K | f = 1)$ in observing a King given a face card was drawn.
 - The entropies $H(S)$ and $H(V, S)$.
 - The mutual information $I(V; S)$ between V and S .
 - The mutual information $I(V; C)$ between the value and colour of a card using the last result and the *data processing inequality*.
4. Recall that for a random variable X , its variance is

$$\text{Var}[X] = E[X^2] - (E[X])^2.$$

Using Jensen’s inequality, show that the variance must always be nonnegative.

5. Let X and Y be independent random variables with possible outcomes $\{0, 1\}$, each having a Bernoulli distribution with parameter $\frac{1}{2}$, i.e.:

$$p(X = 0) = p(X = 1) = \frac{1}{2}$$

- Compute $I(X; Y)$.
 - Let $Z = X + Y$. Compute $I(X; Y | Z)$.
 - Do the above quantities contradict the data processing inequality?
6. Consider a discrete variable X taking on values from the set \mathcal{X} with probability p_i of each state, with $i = 1, \dots, |\mathcal{X}|$. Denote the vector of probabilities by \mathbf{p} . We saw in lectures that the entropy of X satisfies:

$$H(X) \leq \log |\mathcal{X}|,$$

with equality if and only if $p_i = \frac{1}{|\mathcal{X}|}$ for all i , i.e. \mathbf{p} is uniform. Prove the above statement using Gibbs’ inequality, which says

$$\sum_{i=1}^{|\mathcal{X}|} p_i \log_2 \frac{p_i}{q_i} \geq 0$$

for any probability distributions \mathbf{p}, \mathbf{q} over $|\mathcal{X}|$ outcomes, with equality if and only if $\mathbf{p} = \mathbf{q}$.