COMP2610/COMP6261
Tutorial 4 Sample Solutions

Tutorial 4: Entropy and Information

Young Lee and Bob Williamson
Tutors: Debashish Chakraborty and Zakaria Mhammedi

Week 5 (21st - 25th August), Semester 2, 2017

1. Suppose Y is a geometric random variable, Y ~ Geom(y) . i.e., Y has probability function

$$P(Y = y) = p(1-p)^{y-1},\ y = 1, 2, \ldots$$

Determine the mean and variance of the geometric random variable.

Solution.

The expectation of the geometric random variable can be calculated as:

$$\begin{aligned}
E[Y] &= \sum_{y=1}^{\infty} y \cdot P(Y = y) \\
&= \sum_{y=1}^{\infty} y \cdot p(1-p)^{y-1} \\
&= p \sum_{y=1}^{\infty} y(1-p)^{y-1} \\
E[Y] &= p[1 + 2(1-p) + 3(1-p)^2 + \ldots] && (1) \\
(1-p)E[Y] &= [(1-p) + 2(1-p)^2 + 3(1-p)^3 + \ldots] && (2) \\
E[Y] \cdot (1 - (1-p)) &= p[1 + (1-p) + (1-p)^2 + \ldots] && (1) - (2) \\
E[Y] \cdot p &= p \cdot \frac{1}{(1 - (1-p))} && (*) \\
E[Y] &= \frac{1}{p} && (3)
\end{aligned}$$

(*) Here we use the sum to infinity of geometric series, where $|p| < 1$,

$$\sum_{i=1}^{\infty} p^i = \frac{1}{1-p} \tag{4}$$

To calculate the variance, we need to calculate $E[Y^2]$:

$$
\begin{aligned}
E[Y^2] &= \sum_{y=1}^{\infty} y^2 \cdot P(Y=y) \\
&= \sum_{y=1}^{\infty} y^2 \cdot p(1-p)^{y-1} \\
&= \sum_{y=1}^{\infty} (y-1+1)^2 \cdot p(1-p)^{y-1} \\
&= \sum_{y=1}^{\infty} ((y-1)^2 + 2(y-1) + 1) \cdot p \cdot r^{y-1} && \text{let } r = 1-p \\
&= \sum_{z=0}^{\infty} z^2 p r^z + 2\sum_{z=0}^{\infty} z p r^z + \sum_{z=0}^{\infty} p r^z && \text{let } z = y-1 \\
&= r \cdot \sum_{z=0}^{\infty} z^2 p r^{z-1} + 2r \cdot \sum_{z=0}^{\infty} z p r^{z-1} + p \sum_{z=0}^{\infty} r^z \\
&= r \cdot \sum_{z=1}^{\infty} z^2 p r^{z-1} + 2r \cdot \sum_{z=1}^{\infty} z p r^{z-1} + p \cdot \frac{1}{1-(1-p)} && \text{using (4)} \\
E[Y^2] &= r \cdot E[Y^2] + 2r \cdot E[Y] + 1 \\
E[Y^2] &= \frac{1+r}{p^2} \qquad (5)
\end{aligned}
$$

$\therefore$ the variance can be calculated as

$$
\begin{aligned}
Var[Y] &= E[Y^2] - (E[Y])^2 \\
&= \frac{1+r}{p^2} - (\frac{1}{p})^2 && \text{using (5)} \\
&= \frac{r}{p^2} \\
&= \frac{1-p}{p^2} \qquad (6)
\end{aligned}
$$

2. 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:

   (a) The information $h(c = \text{red}, v = \text{K})$ in observing a red King
   (b) The conditional information $h(v = \text{K}|f = 1)$ in observing a King given a face card was drawn.
   (c) The entropies $H(S)$ and $H(V, S)$.
   (d) The mutual information $I(V; S)$ between V and S.
   (e) The mutual information $I(V; C)$ between the value and colour of a card using the last result and the *data processing inequality*.

   Solution.

   (a) $h\left(c = \text{red}, v = \text{K}\right) = \log_2 \frac{1}{P\left(c=\text{red}, v=\text{K}\right)} = \log_2 \frac{1}{1/26} = 4.7004$ bits.
   (b) $h\left(v = \text{K}|f = 1\right) = \log_2 \frac{1}{P\left(v=\text{K}|f=1\right)} = \log_2 \frac{1}{1/3} = 1.585$ bits.
   (c) We have
      i. $H\left(S\right) = \sum_s p\left(s\right) \log_2 \frac{1}{p(s)} = 4 \times \frac{1}{4} \times \log_2 \frac{1}{1/4} = 2$ bits.
      ii. $H(V, S) = \sum_{v,s} p(v, s) \log_2 \frac{1}{p(v,s)} = 52 \times \frac{1}{52} \log_2 \frac{1}{1/52} = 5.7$ bits.
   (d) Since V and S are independent we have $I\left(V; S\right) = 0$ bits.
   (e) Since C is a function of S and by the data processing inequality $I\left(V; C\right) \leq I\left(V; S\right) = 0$. However, mutual information must be nonnegative so we must have $I\left(V; C\right) = 0$ bits.

3. Recall that for a random variable X, its variance is

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

   Using Jensen's inequality, show that the variance must always be non-negative.

   Solution.

   This is a direct application of Jensen's inequality to the convex function $g(x) = x^2$.

4. 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.

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

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

(a) Compute $I(X;Y)$.

(b) Let $Z = X + Y$. Compute $I(X;Y|Z)$.

(c) Do the above quantities contradict the data-processing inequality? Explain your answer.

Solution.

(a) We see that $I(X;Y) = 0$ as $X \perp\!\!\!\perp Y$.

(b) To compute $I(X;Y|Z)$ we apply the definition of conditional mutual information:

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

Now, X is fully determined by Y and Z. In other words, given Y and Z there is only one state of X that is possible, i.e it has probability 1. Therefore the entropy $H(X|Y,Z) = 0$. We have that:

$$I(X;Y|Z) = H(X|Z)$$

To determine this value we look at the distribution $p(X|Z)$, which is computed by considering the following possibilities:

X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	2

Therefore:

$$\mathbf{p}(X|Z=0) = (1,0)$$
$$\mathbf{p}(X|Z=1) = (1/2,1/2)$$
$$\mathbf{p}(X|Z=2) = (0,1)$$

From this, we obtain: $H(X|Z=0) = 0$, $H(X|Z=2) = 0$, $H(X|Z=1) = 1$ bit. Therefore:

$$I(X;Y|Z) = p(Z=1)H(X|Z=1) = (1/2)(1) = 0.5 \text{ bits.}$$

(c) This does not contradict the data-processing inequality (or more specifically the "conditioning on a downstream variable" corollary): the random variables in this example do not form a Markov chain. In fact, Z depends on both X and Y.