EECS 769 Homework 2

Elise McEllhiney

Sept. 20, 2016

0.1 The Value of a Question

0.1.1 Find the decrease in uncertainty $H(X) - H(X|Y)$

$$\begin{aligned} H(X) - H(X|Y) &= H(Y) - H(Y|X) \\ &= H(Y) - 0 \\ &= H(Y) \\ &= H(\alpha) \end{aligned} \tag{1}$$

Apparently any set S with a given α is as good as any other.

0.2 Random Questions

0.2.1 Show $I(X; Q, A) = H(A|Q)$. Interpret.

$$\begin{aligned} I(X; Q, A) &= H(Q, A) - H(Q, A|X) \\ &= H(Q) + H(A|Q) - H(Q|X) - H(A|Q, X) \\ &\stackrel{(a)}{=} H(Q) + H(A|Q) - H(Q) - H(A|Q, X) \\ &\stackrel{(b)}{=} H(Q) + H(A|Q) - H(Q) \\ &= H(A|Q) \end{aligned} \tag{2}$$

(a) Since Q and X are independent $H(Q|X) = H(Q)$
(b) We know A given (Q, X) so $H(A|Q, X) = 0$

Interpretation: The uncertainty in the Answer given the Question, $H(A|Q)$, is the same as the uncertainty in X removed by (Q, A)

0.2.2 Now suppose that two i.i.d. questions Q_1, Q_2 $r(q)$ are asked, eliciting answers A_1 and A_2. Show that two questions are less valuable than twice a single question in the sense that $I(X; Q_1, A_1, Q_2, A_2) \leq 2I(X; Q_1, A_1)$.

$$\begin{aligned} I(X; Q_1, A_1, Q_2, A_2) &= I(X; Q_1) + I(X; A_1|Q_1) + I(X; Q_2|Q_1, A_1) + I(X; A_2|Q_1, A_1, Q_2) \\ &\stackrel{(a)}{=} I(X; A_1|Q_1) + I(X; Q_2|Q_1, A_1) + I(X; A_2|Q_1, A_1, Q_2) \\ &= I(X; A_1|Q_1) + H(Q_2|A_1, Q_1) - H(Q_2|X, A_1, Q_1) + I(X; A_2|Q_1, A_1, Q_2) \\ &\stackrel{(b)}{=} I(X; A_1|Q_1) + H(Q_2) - H(Q_2) + I(X; A_2|Q_1, A_1, Q_2) \\ &= I(X; A_1|Q_1) + I(X; A_2|Q_1, A_1, Q_2) \\ &= I(X; A_1|Q_1) + H(A_2|Q_1, A_1, Q_2) - H(A_2|X, Q_1, A_1, Q_2) \\ &\stackrel{(c)}{=} (X; A_1|Q_1) + H(A_2|Q_1, A_1, Q_2) \\ &\stackrel{(d)}{\leq} (X; A_1|Q_1) + H(A_2|Q_2) \\ &\stackrel{(e)}{=} I(X; A_1|Q_1) + I(X; A_2|Q_2) \\ &= 2I(X; A_1|Q_1) \end{aligned} \tag{3}$$

(a) X and Q_1 are independent thus $I(X;Q_1) = 0$
(b) Q_2 is independent of Q_1, A_1, and X
(c) Given X and Q_2 we can completely determine A_2 so $H(A_2|X,Q_1,A_1,Q_2) = 0$
(d) Entropy never increases with conditioning (e) Proven in 0.2.1

0.3 Entropy of a disjoint mixture

0.3.1 Find $H(X)$ in terms of $H(X_1)$ and $H(X_2)$ and α

$$X = \begin{cases} X_1, & \text{with probability } \alpha \\ X_2, & \text{with probability } 1-\alpha \end{cases}$$

$$f(X) = \begin{cases} 0, & \text{when } X = X_1 \\ 1, & \text{when } X = X_2 \end{cases}$$

$$\begin{aligned} H(X) &= H(X|f(X)) \\ &= H(f(X) + H(X|f(X)) \\ &= H(f(X)) + p(f(X)=0)H(X|f(X)=0) + p(f(X)=1)H(X|f(X)=1) \\ &= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2) \end{aligned} \tag{4}$$

where $H(\alpha) = -\alpha log(\alpha) - (1-\alpha)log(1-\alpha)$

0.3.2 Maximize over α to show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

$$\frac{dH(X))}{d\alpha} = H(X_1) - H(X_2) + log(\frac{1-\alpha}{\alpha})$$

Set the derivative to 0

$$0 = H(X_1) - H(X_2) + log(\frac{1-\alpha}{\alpha})$$

$$-H(X_1) + H(X_2) = log(\frac{1-\alpha}{\alpha})$$

$$2^{-H(X_1)+H(X_2)} = \frac{1-\alpha}{\alpha}$$

$$\alpha = \frac{1}{2^{-H(X_1)+H(X_2)} + 1}$$

$$\alpha = \frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}}$$

$$\begin{aligned} H(X) &= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2) \\ &\leq log(2^{H(X_1)} + 2^{H(X_2)}) \\ 2^{H(X)} &\leq 2^{H(X_1)} + 2^{H(X_2)} \end{aligned} \tag{5}$$

Interpretation: The effective alphabet size of X, a combination of X_1 and X_2, is less than or equal to their effective alphabet sizes

0.3.3 Let X_1 and X_2 be uniformly distributed over their alphabets. What is the maximizing α and the associated $H(X)$

The maximizing α is worked out above.

$$H(X) = log(2^{H(X_1)} + 2^{H(X_2)})$$

The effective alphabet size of two disjoint uniformly distributed random variables is equal to the sum of their effective alphabet sizes

0.4 Bottleneck

0.4.1 Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq log(k)$

$$\begin{aligned} I(X_1; X_3) &\overset{(a)}{\leq} I(X_1; X_2) \\ &= H(X_2) - H(X_2|X_1) \\ &\leq H(X_2) \\ &\leq log(k) \end{aligned} \tag{6}$$

(a) Since we are given that we are working with a Markov chain, we know that mutual information between X_1 and X_3 is less than the mutual information between X_1 and X_2

0.4.2 Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck

$$\begin{aligned} I(X_1; X_3) &\leq log(k) \\ &\leq log(1) = 0 \end{aligned} \tag{7}$$

Since entropy is always nonnegative we know that $I(X_1; X_3) = 0$ and thus for $k = 1$ we know that X_1 and X_3 must be independent

0.5 Relative entropy is not symmetric

0.5.1 Calculate $H(p), H(q), D(p||q)$ and $D(q||p)$

$$H(p) = -\sum p(x)log(p(x)) = \frac{1}{2}log(2) + \frac{1}{4}log(4) + \frac{1}{4}log(4) = 1.5bits$$

$$H(q) = -\sum q(x)log(q(x)) = \frac{1}{3}log(3) + \frac{1}{3}log(3) + \frac{1}{3}log(3) = 1.585bits$$

$$D(p||q) = \sum p(x)log\frac{p(x)}{q(x)} = \frac{1}{2}log(\frac{3}{2}) + \frac{1}{4}log(\frac{3}{4}) + \frac{1}{4}log(\frac{3}{4}) = 0.085$$

$$D(q||p) = \sum q(x)log\frac{q(x)}{p(x)} = \frac{1}{3}log(\frac{2}{3}) + \frac{1}{3}log(\frac{4}{3}) + \frac{1}{3}log(\frac{4}{3}) = 0.082$$

0.5.2 Verify that in this case $D(p||q) \neq D(q||p)$

As we can see from the above equations $D(p||q) \neq D(q||p)$ since $0.85 \neq 0.082$

0.6 Conditional mutual information

0.6.1 Find the mutual information

Since we know that we must have an even number of ones, we know that the n^{th} digit is determined by the previous $n-1$ digits. Since we know that any sequence with an even number of digits is equally likely, we know that any $n-1$ or fewer of the digits are independent.
For $k \leq n-1$ we know that:

$$I(X_{k-1}; X_k | X_1, X_2, ..., X_{k-2}) = 0$$

However, if we know $n-1$ digits, the last digit is deterministic.

$$I(X_{n-1}; X_n | X_1, X_2, ..., X_{n-2}) = H(X_n | X_1, X_2, ..., X_{n-2}) - H(X_n | X_1, X_2, ..., X_{n-1}) = 1bit$$

0.7 Conditional mutual information vs unconditional mutual information

0.7.1 Give examples of joint random variables such that X, Y, and Z such that:

(a)$(I(X;Y|Z) < I(X;Y)$

X is a uniform binary random variable where $Y = X$ and $Z = Y$

$$I(X;Y) = H(X) - H(X|Y) = H(X) = 1$$
$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = 0$$

(b)$I(X;Y|Z) > I(X;Y)$

X and Y are independent uniform binary random variables. $Z = X + Y$

$$I(X;Y) = H(X) - H(X|Y) = 0$$
$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|Z) = 1/2$$

0.8 Entropy of initial conditions

0.8.1 Prove that $H(X_0|X_n)$ is non-decreasing with n for any Markov chain

The data processing inequality states that for any Markov chain, $I(X;Y) \geq I(X;Z)$

$$I(X_0; X_{n-1}) \geq I(X_0; X_n)$$
$$H(X_0) - H(X_0|X_{n-1} \geq H(X_0) - H(X_0|X_n)$$

Thus, $H(X_0|X_n)$ is non-decreasing as n increases for any Markov chain

0.9 A metric

$$\rho(x,y) \geq 0$$
$$\rho(x,y) = \rho(y,x)$$
$$\rho(x,y) = 0 iff x = y$$
$$\rho(x,y) + \rho(y,z) \geq \rho(x,z)$$

0.9.1 Show that $\rho(X,Y) = H(X|Y) + H(Y|X)$ satisfies the first, second and fourth properties above. If we say that $X = Y$ if there is a one-to-one function mapping X to Y, then the third property is also satisfied, and $\rho(X,Y)$ is a metric

$$\rho(X,Y) = H(X|Y) + H(Y|X)$$

1) $\rho(X,Y) \geq 0$ since entropy is always nonnegative.
2) $\rho(X,Y) = \rho(Y,X)$ since $H(X|Y) + H(Y|X) = H(Y|X) + H(X|Y)$
3) We know that $H(X|Y) = 0$ iff $X = g(Y)$ and that $H(Y|X) = 0$ iff $Y = g(X)$. From this we know that $\rho(X,Y) = 0$ iff $Y = g(X)$ and $X = g(Y)$. Thus, we know that $p(X,Y) = 0$ iff there is a one-to-one mapping from X to Y
4)$\rho(x,y) + \rho(y,z) \geq \rho(x,z)$ since:

$$\begin{aligned} H(X|Y) + H(Y|Z) &\geq H(X|Y,Z) + H(Y|Z) \\ &= H(X,Y|Z) \\ &= H(X|Z) + H(Y|X,Z) \\ &\geq H(X|Z) \end{aligned} \tag{8}$$

0.9.2 Verify that $\rho(X,Y)$ can also be expressed as

$$\begin{aligned} \rho(X,Y) &= H(X) + H(Y) - 2I(X;Y) \\ &= H(X,Y) - I(X;Y) \\ &= 2H(X,Y) - H(X) - H(Y) \end{aligned} \tag{9}$$

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

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

Therefore

$$\rho(X,Y) = H(X) - I(X;Y) + H(Y) - I(X;Y) = H(X) + H(Y) - 2I(X;Y)$$

$$H(X,Y) = H(X) + H(Y) - I(X,Y)$$

Thus, combining the previous two equations, we get:

$$\rho(X,Y) = H(X,Y) - I(X,Y)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Combining the previous two equations we get:

$$\rho(X,Y) = 2H(X,Y) - H(X) - H(Y)$$

0.10 Run length coding

0.10.1 Show all equalities and inequalities, and give simple bounds on all differences

We know that $H(R) \leq H(X)$ since we know that the run lengths, R, are a function of $X_1, X_2, ..., X_n$ Knowing both the run lengths and any X_i in the sequence, we can determine the entire sequence $X_1, X_2, ..., X_n$

$$\begin{aligned} H(X_1, X_2, ..., X_n) &= H(X_i, R) \\ &= H(R) + H(X_i|R) \\ &\leq H(R) + H(X_i) \\ &\leq H(R) + log(2) \\ &= H(R) + 1 \end{aligned} \tag{10}$$

0.11 Pure randomness and bent coins

0.11.1 Give reasons for the following inequalities

$$\begin{aligned} nH(p) &\overset{(a)}{=} H(X_1, ..., X_n) \\ &\overset{(b)}{\geq} H(Z_1, Z_2, ..., Z_K, K) \\ &\overset{(c)}{=} H(K) + H(Z_1, ..., Z_K|K) \\ &\overset{(d)}{=} H(K) + E(K) \\ &\overset{(e)}{\geq} EK \end{aligned} \tag{11}$$

(a) $X_i = 1$ with the probability p. Also, we know that $X_1, X_2, ..., X_n$ are independent and identically distributed. Thus, we can conclude that $nH(p) = H(X_1, X_2, ..., X_n)$. In other words, the entropy of a sequence is the product of the entropies of the random variables
(b) We know that $(Z_1, Z_2, ..., Z_K) = f(X_1, X_2, ..., X_n)$. Since we know that the entropy of a random variable is always greater than or equal to the entropy of a function of that random variable, we know that: $H(X_1, X_2, ..., X_n) \geq H(Z_1, Z_2, ..., Z_K)$. Since K is a function of $Z_1, Z_2, ..., Z_K$ we know that $H(K|Z_1, Z_2, ..., Z_K) = 0$. Thus:

$$H(Z_1, Z_2, ..., Z_K, K) = H(Z_1, Z_2, ..., Z_K) + H(K|Z_1, Z_2, ..., Z_K) = H(Z_1, Z_2, ..., Z_K)$$

(c)Chain rule
(d)Since $Z_1, Z_2, ..., Z_K$ are random bits with $H(Z_i) = 1bit$.
(e)Entropy is non-negative
Thus no more than $nH(p)$ fair coin tosses can be derived from $(X_1, ..., X_n)$, on the average.

0.11.2 Exhibit a good map f on sequences of length 4

Since it is an unfair coin and we don't know p there are not very many ways to simulate a sequence of fair coin-flips. However, even without knowing p we know that the flips are i.i.d. so we know that a series of 4 flips having a given number of heads occurring is equally likely.

$0000 \to \mathrm{NULL}$					
$0001 \to 00$	$0010 \to 01$	$0100 \to 10$	$1000 \to 11$		
$0011 \to 00$	$0101 \to 01$	$0110 \to 10$	$1010 \to 11$	$1001 \to 0$	$1100 \to 1$
$0111 \to 00$	$1011 \to 01$	$1101 \to 10$	$0111 \to 11$		
$1111 \to \mathrm{NULL}$					