# Homework 2

**Biophysics of Biomolecules**

*April 13th, 2021*

Clayton Seitz

**Problem 0.1.** *Two teams A and B play a series of up to 5 games, in which the team to win 3 games wins the series. Let X be a random variable which is a sequence of letters corresponding to the winners of each of the games played - possible values for X then include AAA, ABBAB etc. Let Y be the number of games played (the teams play till the series winner is decided). Calculate* $H(X), H(Y), H(X|Y), H(Y|X)$ *and* $I(X;Y)$*. Assume both teams are equally likely to win each game independent of any previous games.*

**Solution**. The series can be 3, 4, or 5 games long. There are a total of $2^5 = 32$ possible 5-bit strings but we need to truncate strings after one team has won. Starting with 32 leaves, any nodes below strings 000 or 111 in the binary tree should be deleted. This trims $2^5 \cdot \frac{2}{2^3} = 8$ leaves down to 2. Also, there are 6 ways the series can end after 4 games, so we trim $2^5(\frac{6}{2^4}) = 12$ leaves to 6. We have removed a total of 12 leaves leaving us with 20 possible leaves (series).

$$H(X) = \sum_X P(X)\frac{1}{\log P(X)} = \log 20 \approx 4.3\text{bits}$$

where $\log x = \log_2 x$. There are 2 ways the series can end after 3 games, 6 ways it can end after 4 games, and 12 ways it can end after 5 games. If $Y$ is the number of games played,

$$H(Y) = \sum_Y P(Y)\frac{1}{\log P(Y)} = \frac{1}{10}\log 10 + \frac{3}{10}\log\frac{10}{3} + \frac{6}{10}\log\frac{5}{3} \approx 1.3\text{bits}$$

The conditional entropy $H(X|Y)$ is given by

$$H(X|Y) = \sum_y P(y)H(X|Y=y) = \frac{1}{10}\cdot 1 + \frac{3}{10}\cdot\log 6 + \frac{6}{10}\cdot\log 12 \approx 3\text{bits}$$

We see that conditioning on $Y$ reduced the entropy in $X$. The conditional entropy $H(Y|X)$ is

$$H(Y|X) = \sum_x P(x)H(Y|X=x) = 0\text{bits}$$

since if we are given the series, it is certain how many games were played. Finally the mutual information $I(X;Y)$ is

$$I(X;Y) = H(X) - H(X|Y) = 1.3\text{bits}$$

■

**Problem 0.2.** *Lost in transmission*

**Solution**.

$$\begin{aligned} I(X_n;X_1) &= H(x_n) - H(x_n|x_1) \\ &= \sum_y P(x_1=y) \sum_x P(x_n=x|x_1=y) \log \frac{1}{P(x_n=x|x_n=y)} \end{aligned}$$

where $x, y \to 0, 1$. We can compute $P(x_n = x_1)$ by noticing that this will only occur if there is an even number of bit flips. If $\epsilon$ is the probability of a flip which we can define as a "success" and each trial is independent, then $P(x_n = x_1)$ is a sum over the binomial distribution $B(n,k)$ for even values of $k$. Define $P(x_n = x_1)$ as

$$\Omega = 1 - \frac{1}{2}[1 + (1-2\epsilon)^n]$$

From which we can compute the entropy

$$\begin{aligned} H(x_n|x_1) &= \Omega \log \frac{1}{\Omega} + (1-\Omega) \log \frac{1}{1-\Omega} \\ &= H_b(\Omega) \end{aligned}$$

and because $H(x_n) = 1$, the mutual information is given by

$$I(X_n;X_1) = 1 - \Omega \log \frac{1}{\Omega} + (1-\Omega) \log \frac{1}{1-\Omega}$$

As a sanity check, we can show that for $n = 1$ this takes the form of a binary symmetric channel and when $\epsilon = 0$ or $\epsilon = 1$ we have $I(x_1; x_n) = 1$.

■

**Problem 0.3.** *Prove the following basic identities about the quantities we have studied so far.*

**Solution**.

$$\begin{aligned} D(P||Q) &= \sum P(x) \log \frac{P(x)}{Q(x)} && (1)\\ &= \sum P(x) \log |\chi| P(x) && (2)\\ &= \sum P(x) \log P(x) + \sum P(x) \log |\chi| && (3)\\ &= \log |\chi| - H(x) && (4) \end{aligned}$$

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (5)\\ &= \sum P(x) \log \frac{1}{P(x)} + \sum P(x,y) \log \frac{P(x,y)}{P(y)} && (6)\\ &= \sum P(x,y) \log \frac{P(x,y)}{P(x)P(y)} && (7)\\ &= D(P(x,y)||P(x)P(y)) && (8) \end{aligned}$$

■

**Problem 0.4.** *Expand our definition for mutual information between three random variables $X, Y, Z$.*

**Solution**.

$$\begin{aligned} I(X;Y;Z) &= I(X;Y) - I(X;Y|Z)\\ &= [H(X) + H(Y) - H(XY)] - [H(XZ) + H(YZ) - H(XYZ) - H(Z)]\\ &= H(X) + H(Y) + H(Z) - H(XY) - H(XZ) - H(YZ) + H(XYZ) \end{aligned}$$

■

We need to design $X, Y, Z$ such that $I(X;Y|Z) > I(X;Y)$. Lets consider a scenario where $I(X;Z) = I(Y;Z) = 0$ but together $Y, Z$ carry information about $X$.

$$(X,Y,Z) = \begin{cases} 011 & w.p\ 1/4 \\ 000 & w.p\ 1/4 \\ 101 & w.p\ 1/4 \\ 110 & w.p\ 1/4 \end{cases}$$

It is necessary to show that $I(X;Y|Z) > 0$. Indeed,

$$\begin{aligned} I(X;Y|Z) &= \mathbf{E}_z[I(X|Z=z;Y|Z=z)] \\ &= \frac{1}{2}I(X|Z=0;Y|Z=0) + \frac{1}{2}I(X|Z=1;Y|Z=1) \\ &= \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1 \end{aligned}$$

which means that $I(X;Y;Z) < 0$.

**Problem 0.5.** *Prove that* $\mathbf{E}_x||P(Y|X=x) - P(Y)||_1 \leq \sqrt{2\ln 2\ I(X;Y)}$

**Solution**. We can use Pinsker's inequality

$$||P(Y|X) - P(Y)||_1 \leq \sqrt{2\ln 2\ D_{KL}(P(Y|X)||P(Y))}$$

$$\begin{aligned} \mathbf{E}_x||P(Y|X=x) - P(Y)||_1 &\leq \mathbf{E}_x\sqrt{2\ln 2\ D_{KL}(P(Y|X=x)||P(Y))} \\ &\leq \sqrt{2\ln 2 \cdot \mathbf{E}_x\ D_{KL}(P(Y|X=x)||P(Y))} \\ &= \sqrt{2\ln 2\ I(X;Y)} \end{aligned}$$

■

**Problem 0.6.** *Prove that*

$$\sum(\frac{\sqrt{5}-1}{2})^{e_i} \leq 1$$

**Solution**. Let the sum of energies of all possible codes in a tree of depth $l_{max}$ be $E$. We need to show that the fraction of $E$ we delete when we select a code of energy $e_i$ is equal to $(\frac{\sqrt{5}-1}{2})^{e_i}$.

■