Problem Set 1

2.1 Coin Flips

(a)

The inputs to the random variable X mentioned in the question are in the following form:

H, TH, TTH, TTTH, ...

These inputs represent sequences of coin tosses performed until an H is observed. X outputs the length of a given such sequence. For instance, the outputs corresponding to the above sequences are:

1, 2, 3, 4, ...

Because a sequence of coin tosses are independent, the probability for this random variable can be defined as:

$$P(X = x) = \frac{1}{2^x} \tag{1}$$

where x is the number of tosses performed to get an H. Remember that for a fair coin, the probability of getting a head is the same as the probability of getting a tail i.e. $\frac{1}{2}$. And as the tosses are independent, you just multiply the probabilites of each individual outcome to get the probability of a whole sequence.

As the entropy is to be calculated in bits, logarithm of base 2 will be used. Now, we know:

$$\begin{aligned}
H(X) &= -\sum_{x=1}^{\infty} P(X = x) \cdot log_2(P(X = x)) \\
&= -\sum_{x=1}^{\infty} \frac{1}{2^x} \cdot log_2(\frac{1}{2^x}) && [(1)] \\
&= \sum_{x=1}^{\infty} \frac{1}{2^x} \cdot log_2(2^x) && [log(\frac{1}{x}) = -log(x)] \\
&= \sum_{x=1}^{\infty} \frac{x}{2^x} \cdot log_2(2) && [log(a^x) = xlog(a)] \\
&= \sum_{x=1}^{\infty} \frac{x}{2^x} && [2^1 = 2 \implies log_2(2) = 1] \\
&= 2
\end{aligned}$$

You can show that the series $\sum_{x=1}^{\infty} \frac{x}{2^x}$ converges to 2. Look at the link here.

(b)

TODO

2.2 Entropy of functions

(a)

2^x gives a unique output for a unique input $x \in X$, that is, 2^x is one to one. Therefore, the cardinality of Y will be the same as X. This also means that for any $y \in Y$, $P(Y = y)$ will be the same as $P(X = x)$, where $x \in X$ is its corresponding input. Remember that in the equation of entropy, you never once directly use the value of $y \in Y$, only the probability $P(Y = y)$. Therefore, given all these facts, $H(Y) = H(X)$, trivially.

(b)

This is trickier as $cos(x)$ can give the same output for unique inputs $x \in X$. For instance, $cos(0°) = cos(360°) = 1$. You can select a finite domain for X such that $cos(X)$ is one to one, like from $0°$ to $90°$. In this case, by the same argument as **(a)**, $H(Y) = H(X)$. But this is obviously not guaranteed.

Let's assume that unique $x_1 \in X$ and $x_2 \in X$ both give the same $y \in Y$. The probability of y here will be:

$$P(Y = y) = P(X = x_1) + P(X = x_2) \tag{2}$$

Now, if you think about it, $H(Y)$ is going to contain the following term in its summation:

$$P(Y = y) \cdot log(P(Y = y)) = (P(X = x_1) + P(X = x_2)) \cdot log(P(X = x_1) + P(X = x_2)) \tag{3}$$

But not this term:

$$P(X = x_1) \cdot log(P(X = x_1)) + P(X = x_2) \cdot log(P(X = x_2)) \tag{4}$$

$H(X)$ will have (4) but not (3), and $H(Y)$ will have (3) but not (4). Let's compare (3) and (4).

$log(P(X = x_1))$ and $log(P(X = x_2))$ are both greater than $log(P(X = x_1) + P(X = x_2))$ in magnitude, as both $P(X = x_1)$ and $P(X = x_2)$ are individually less than their sum, and logarithm is monotonically increasing while giving negative outputs between 0 and 1 (remember that probabilities are between 0 and 1). This implies that, for magnitudes of the logarithms, (4) is greater than (3). This is actually enough to imply that $H(X) > H(Y)$. Because it is trivial to show that entropy can also be written as:

$$H(X) = \sum_{x \in X} P(x) \cdot |log(P(x))| \tag{5}$$

Basically, $H(X)$ and $H(Y)$ have corresponding terms, but the term in $H(X)$ is greater than the term in $H(Y)$, therefore, $H(X) > H(Y)$, all else being the same. If all else isn't the same, that is because there are more than one $y \in Y$ that have more than one corresponding inputs in X. But this would only support $H(X) > H(Y)$. This argument also trivially scales for the case when there are more than two inputs that give the same output.