Problem 3.1.

Paolo is an avid coin collector. On his shelf he has two jars filled with different types of coins. The first jar contains 300 fair coins that land on heads or tails with equal probability. The second jar contains 100 carefully rigged coins that land on heads with probability $\frac{2}{3}$, and on tails with probability $\frac{1}{3}$. One day when Paolo came home from work his shelf was broken and all of his coins were spilled on the floor. He collected them, and none of the coins were missing, but he could not distinguish the coins anymore so he put all of them in a new, plastic jar. He then extracted a random coin from the plastic jar and threw it twice.
Let $S_0 \in \{F, R\}$ denote the type of coin Paolo chose (F = fair, R = rigged), and $S = (S_1, S_2)$ be the outcomes of the first and second throws respectively, where $S_1, S_2 \in \{H, T\}$.

1. Compute $p_{S_1|S_0}$, $p_{S_2|S_0}$ and $p_{S|S_0}$.
2. Compute $H(S_0)$, $H(S_1)$, $H(S_2)$ and $H(S)$.
   Are S_1 and S_2 independent?
3. Compute $H(S_1|S_0)$, $H(S_2|S_0)$ and $H(S|S_0)$ using the definition of conditional entropy. Compare the values obtained with those computed in part 2 and explain any differences.
4. Using entropies, can you say if S_1 and S_2 are independent if you observe S_0 first?
5. Compute $H(S_0, S_1, S_2)$ and $H(S_0|S)$ only using the chain rule of conditional entropy.

Problem 3.2.

Consider a random variable X taking values in $\mathcal{X} = \{-3, -2, -1, 0, 1, 2, 3\}$. Denote with P_X its distribution. We have that $P_X(0) = 1/2$ while $P_X(x) = 1/12$ for $x \neq 0$.

1. Compute the entropy of $H(X)$.

Let now $Y = 2X + 2$.

3. Compute $H(Y)$;
4. Compute $H(Y|X)$;
5. Compute $H(X|Y)$;
6. Compute $H(X, Y)$:
   (a) using the chain rule for entropy ($H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$);
   (b) using (after computing) $P_{XY}(x, y)$ for all pairs x, y;

Let $Z = X^2$.

7. Compute $H(Z)$;
8. Compute $H(Z|X)$;
9. Compute $H(X|Z)$;
10. Compute $H(X, Z)$:
    (a) using the chain rule for entropy ($H(X, Z) = H(X) + H(Z|X) = H(Z) + H(X|Z)$);
    (b) using (after computing) $P_{XZ}(x, z)$ for all pairs x, z;

These two random variables Y, Z are specific cases of the more general setting $W = f(X)$ where $f : \mathcal{X} \to \mathcal{W}$ is a deterministic function (mapping each element in the alphabet of

X to only one element in $\mathcal{W}$ with probability 1, as opposed to random functions that can associate to an element $x \in \mathcal{X}$ two or more elements in $\mathcal{W}$ with a certain probability). What can you deduce from the computations above? What is the difference between Y and Z and the respective entropies $H(Y), H(Z)$ when compared with $H(X)$?

Problem 3.3.

Up until now we have seen 2 special code constructions: Shannon-Fano coding and Huffman coding. Shannon-Fano coding is a simple construction for which $L(S, \Gamma_{SF})$ has a good upper bound, but it is not optimal. On the other hand, Huffman codes are optimal. Does there exist another non-optimal code that performs better than a Shannon-Fano code? We try to answer this question experimentally.

Let S be a source distributed as follows:

S	a	b	c	d	e
p_S	$\frac{10}{52}$	$\frac{9}{52}$	$\frac{8}{52}$	$\frac{5}{52}$	$\frac{20}{52}$

1. Compute $H(S)$ and $L(S, \Gamma_{SF})$ of a binary Shannon-Fano code for source S.
2. Consider the following binary Huffman code H for source S:

Source Symbol	H
a	111
b	110
c	101
d	100
e	0

   Draw its decoding tree and compute $L(S, \Gamma_H)$.
3. Let X_i denote the i-th bit of a Huffman codeword (code H). Find the distributions $p_{X_1}(\cdot)$, $p_{X_2|X_1}(\cdot|1)$, $p_{X_3|X_1,X_2}(\cdot|1,0)$, $p_{X_3|X_1,X_2}(\cdot|1,1)$ and their respective entropies. What do you notice about the values of these entropies ?
4. Fano coding is another suboptimal coding scheme that generally produces a shorter average codeword length than Shannon-Fano coding.
   Below lies the detailed Fano coding procedure:
   (a) Arrange the source probabilities in decreasing order. Let p' be the list obtained.
   (b) Without changing the order, split p' into two halves $p' = (p'_L, p'_R)$ such that the absolute value of the difference between the sum over the elements of p'_L and the sum over the elements of p'_R is as small as possible. Thus, the two halves p'_L and p'_R are as equiprobable as possible.
   (c) Set the first bit of p'_L to 0 and the first bit of p'_R to 1.
   (d) Repeat steps (b) and (c) recursively for each subset until termination.

   Find a Fano code F for source S and draw its decoding tree. Compare $H(S)$, $L(S, \Gamma_H)$, $L(S, \Gamma_{SF})$, and $L(S, \Gamma_F)$.