Problem 1. a) First define a new random variable θ as follows:

$$\theta = f(X) = \begin{cases} 1 & \text{when } X = X_1, \\ 2 & \text{when } X = X_2, \end{cases}$$

Then we will have $H(X, \theta) = H(X) + H(\theta|X) = H(X)$ since θ is a function of X.

On the other hand we have $H(X, \theta) = H(\theta) + H(X|\theta)$. We also know that $H(X|\theta) = p(\theta = 1)H(X|\theta = 1) + p(\theta = 2)H(X|\theta = 2)$. Now from the assumption we have $p(\theta = 1) = \alpha$ and $p(\theta = 2) = 1 - \alpha$. Finally with substitution we have: $H(X) = H(\theta) + \alpha H(X_1) + (1 - \alpha)H(X_2) = H(\alpha) + \alpha H(X_1) + (1 - \alpha)H(X_2)$.

b) we differentiate the above expression with respect to α to find the maxima. We have:

$$\begin{aligned} \frac{dH(X)}{d\alpha} &= \frac{d(-\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) + \alpha Y_1 + (1 - \alpha)Y_2)}{d\alpha} \\ &= -\log \alpha + \log(1 - \alpha) + Y_1 - Y_2 \end{aligned}$$

where $Y_i = H(X_i)$. Now, in order to find the maxima of $H(X)$ we must solve the equation $\frac{dH(X)}{d\alpha} = 0$. If we solve this equation for α we can easily see that $\alpha = \frac{2^{Y_2}}{2^{Y_1}+2^{Y_2}}$ and therefore $1-\alpha = \frac{2^{Y_1}}{2^{Y_1}+2^{Y_2}}$. Notice that with this α $H(X)$ is maximized. Therefore in general we have the following inequality for $H(X)$:

$$H(X) \leq -\beta \log \beta - (1 - \beta) \log(1 - \beta) + \beta Y_1 + (1 - \beta)Y_2)$$

Where $\beta = \frac{2^{Y_1}}{2^{Y_1}+2^{Y_2}}$. Equivalently we have:

$$2^{H(X)} \leq 2^{-\beta \log \beta - (1-\beta) \log(1-\beta) + \beta Y_1 + (1-\beta)Y_2)} \tag{1}$$

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

In fact the equality (2) can be obtained using simple calculation an it is straight forward.

Problem 2. a) The number of 100-bit binary sequences with three or fewer ones is:

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 166751.$$

So The required codeword length is $\log_2 166751 = 18$ bits.

b) The probability that a 100-bit sequence has three or fewer ones is equal to:

$$\sum_{i=0}^{3} \binom{100}{i} (0.015)^i (0.985)^{100-i} = 0.935784065.$$

Thus, the probability that the sequence which is generated cannot be encoded is $1 - 0.935784065 = 0.064215935$.

c) In the case of a random variable S_n that is the sum of n i.i.d. random variables $X_1; X_2; \ldots; X_n$, Chebyshev's inequality states:

$$P(|S_N - n\mu| \geq \epsilon) \leq \frac{n\sigma^2}{\epsilon^2}.$$

where μ and σ^2 are the mean and variance of X_i. In this problem, $n = 100, \mu = 0.015$ and $\sigma^2 = (0.015)0.985$. Note $S_{100} \geq 4$ if and only if $|S_{100} - 100(0.015)| \geq 2.5$, so we should choose $\epsilon = 2.5$. Then, $P(S_{100} \geq 4) \leq \frac{100\times 0.015\times 0.985}{2.5^2}$.

Problem 3. a) $H(X|Y) = H(Z + Y|Y) = H(Z|Y)$. Furthermore, since conditioning decreases entropy, $H(Z|Y) \leq H(Z)$ and thus $H(X|Y) \leq H(Z)$

b) $H(X|Y) = H(Z)$ if and only if $H(Z|Y) = H(Z)$. That is Z and Y are independent.

c) We can instead, prove that $I(U;W) + I(U;T) \leq I(U;V) + I(W;T)$. By adding the term $I(U;T|W)$ to both sides, it suffices to show that $I(U;T|W) + I(U;W) + I(U;T) \leq I(U;V) + I(W;T) + I(U;T|W)$

By using chain rule, we have that $I(U;T|W)+I(U;W) = I(U;T,W)$ at the left hand side, and $I(U;T|W) + I(W;T) = I(U,W;T)$ at the right hand side. Thus it suffices to show that $I(U;T,W) + I(U;T) \leq I(U;V) + I(U,W;T)$. From the Markov chain $U \leftrightarrow V \leftrightarrow (W,T)$, $I(U;W,T) \leq I(U;V)$. Furthermore, $I(U;T) \leq I(U,W;T) = I(U;T) + I(W;T|U)$ since $I(W;T|U) \geq 0$. This concludes the solution

Problem 4.

- First we compute $H(X)$. Notice that we can partition all the possible values of X into 4 groups. The first group consists of $NNNN$ and $FFFF$. The second group consists of all the strings of N and F of length 5 so that four symbols are identical and the remaining one is different and also it is not the last one. One can easily observe that there are $2 \times 4 = 8$ possibilities in this group. The third and the fourth groups are defined similarly. (The third group consists of possible strings of length 6 and the fourth group consists of the possible strings of length 7). One can compute the sizes of the third and the fourth group. In fact the third group contains $2 \times \binom{5}{2} = 20$ and the fourth group contains $2 \times \binom{6}{3} = 40$ strings. Since both player are equally matched and the games are independent therefore the probability of each string in the $i-th$ group is equal to 2^{-i-3}. (for example the probability of the event $X = FNNFFF$ is equal to 2^{-6}. Using this information we can easily compute $H(X)$. In fact we can say that:

$$H(X) = 2 \times (2^{-4} \times 4) + 8 \times (2^{-5} \times 5) + 20 \times (2^{-6} \times 6) + 40 \times (2^{-7} \times 7)$$

- Next we compute $H(Y)$. As we saw in the previous part, the first group contains 2 elements each of which with probability 2^{-4}. So, the probability that $Y = 4$ is equal to $2 \times 2^{-4} = 1/8$. Similarly we can find out the probability of the other values of Y. In fact we have: $p(Y = 5) = 1/4, p(Y = 6) = 5/16$ and $p(Y = 7) = 5/16$. So we have $H(Y) = 3/8 + 1/2 + 5/16 \log(16/5) + 5/16 \log(16/5)$

- The next quantity we can easily find is $H(Y|X)$. Notice that if X is given then Y is completely determined. So $H(Y|X) = 0$

- For the final quantity we use the equality $H(X) + H(Y|X) = H(Y) + H(X|Y)$. we already found three of the four. Therefore we can find the fourth quantity.

Problem 5. Notice that this inequality can be also written as $n(H(X)-\epsilon)-1 \leq \log |B^n \cap A^n_{(\epsilon)}| \leq n(H(X)+\epsilon)$. or equivalently

$$\frac{1}{2} 2^{n(H(X)-\epsilon)} \leq |B^n \cap A^n_{(\epsilon)}| \leq 2^{n(H(X)+\epsilon)}.$$

when n is large enough. First we prove the right hand side inequality. Namely, we show that if n is large enough then $|B^n \cap A^n_{(\epsilon)}| \leq 2^{n(H(X)+\epsilon)}$ But notice that $|B^n \cap A^n_{(\epsilon)}| \leq |A^n_{(\epsilon)}| \leq 2^{n(H(X)+\epsilon)}$.

For the other inequality we argue as follows. By the weak law of large numbers we know that $\frac{1}{n}\sum_{i=1}^n X_i$ approaches to $E[X]$ in probability.This means that for every $\delta > 0$, $p(\{x^n \in \mathcal{X}^n : |\frac{1}{n}\sum_{i=1}^n X_i - E[X]| > \delta\})$ goes to zero, as n goes to infinity. In the other words, $p(x^n \in B^n)$ goes to 1 as n goes to infinity. Therefore we can conclude that if n is larger than a constant number N_1 which depends on δ then $p(x^n \in B^n) \geq 1 - \frac{\delta}{2}$. Similarly, if $n > N_2$ for some constant N_2 which depends on δ then $p(x^n \in A^n_{(\epsilon)}) \geq 1 - \frac{\delta}{2}$. Then we use the following equation:

$$p(x^n \in B^n) + p(x^n \in A^n_{(\epsilon)}) = p(x^n \in B^n \cap A^n_{(\epsilon)}) + p(x^n \in B^n \cup A^n_{(\epsilon)})$$

Using the previous inequalities about $p(x^n \in B^n)$ and $p(x^n \in A^n_{(\epsilon)})$ and also the fact that $p(x^n \in B^n \cup A^n_{(\epsilon)}) \leq 1$, we have:

$$2 - \delta \leq 1 + p(x^n \in B^n \cap A^n_{(\epsilon)})$$

and therefore $p(x^n \in B^n \cap A^n_{(\epsilon)}) \geq 1 - \delta$, provided that $n \geq \max\{N_1, N_2\}$.

Now, we try to find a lower bound for $|B^n \cap A^n_{(\epsilon)}|$. Notice that each element of the set $B^n \cap A^n_{(\epsilon)}$ is in particular an element of $A^n_{(\epsilon)}$. Therefore each element of $B^n \cap A^n_{(\epsilon)}$ has probability at most $2^{-n(H(X)-\epsilon)}$. Therefore $p(x^n \in B^n \cap A^n_{(\epsilon)}) \leq |B^n \cap A^n_{(\epsilon)}| \times 2^{-n(H(X)-\epsilon)}$ Combining the inequalities for the lower bound and upper bound for $p(x^n \in B^n \cap A^n_{(\epsilon)})$ we have :

$$1 - \delta \leq p(x^n \in B^n \cap A^n_{(\epsilon)}) \leq |B^n \cap A^n_{(\epsilon)}| \times 2^{-n(H(X)-\epsilon)}$$

Thus $|B^n \cap A^n_{(\epsilon)}| \geq (1-\delta) \times 2^{n(H(X)-\epsilon)}$, provided that $n \geq \max\{N_1, N_2\}$. Since this inequality holds for every positive δ we can take $\delta = 1/2$ and then the left hand side inequality follows.