Please provide detailed answers.

1. (2 points) Consider random variables \( X \sim Ber(0.1) \) and \( Y = X + Z \) where \( Z \sim Ber(0.3) \) is independent of \( X \).
   a. Find the best estimator of \( X \) given \( Y \) and its probability of error.
   b. Compare the result at the previous point with Fano’s inequality.

2. (1 point) Consider a stationary process \( X_1, X_2, \ldots \). Prove that \( H(X_i | X_1, \ldots, X_{i-1}) \) is non-increasing with \( i \).

3. (3 points) We are given two pmfs \( p(x) = (1/12, 1/2, 1/4, 1/6) \) and \( q(x) = (0, 1/2, 1/4, 1/4) \).
   a. Assume that \( q(x) \) is the correct pmf of a memoryless source, but a Huffman code \( C \) is constructed using the wrong pmf \( p(x) \). What is the redundancy \( L(C) - H(X) \) of this code?
   b. Compare the result above with the redundancy of a Shannon code constructed using the wrong pmf \( p(x) \). Connect this result to the KL divergence between \( p(x) \) and \( q(x) \).
   c. Assume now that \( p(x) \) is the correct pmf of a memoryless source, but a Huffman code is constructed using the wrong pmf \( q(x) \). What can you say about the resulting code?

4. (2 points) We want to generate a random variable \( Y \sim p(x) = (1/2, 1/4, 1/4) \). To this end, we have available a fair coin that we can toss independently multiple times, i.e., an iid sequence of variables \( X_i \sim Ber(0.5) \). Find a way to generate \( Y \) from multiple tosses of the coin. Show that this scheme requires on average a number of coin tosses equal to \( H(Y) \).

5. (1 point) Show that an arithmetic code, which has \( l(x) = \lceil -\log_2 p(x) \rceil + 1 \) and \( c(x) = \lceil \bar{F}(x) \rceil_{l(x)} \), is prefix free.

6. (2 points) A Ber\((p)\) memoryless source is compressed using a fixed-to-variable Shannon code that operates over blocks of size \( k \). Note that the per-symbol length \( l(X^k)/k \) of the resulting codeword is a random variable. You can approximate \( l(X^k) \) with the ideal codeword length in order to simplify the problem.
   a. As \( k \) grows, what happens to \( l(X^k)/k \)?
   b. We wish to have a variance of \( l(X^k)/k \) smaller than 0.1. How large should \( k \) be if \( p = 0.1 \)?