Exercise Set #4

1. Consider a source that produces an i.i.d. sequence of symbols from the alphabet \( \{A, B, C, D, E, F, G\} \) with probabilities 0.22, 0.35, 0.15, 0.09, 0.09, 0.05, and 0.05, respectively. Find the binary Shannon–Fano code and compute the average number of code symbols per source symbol.

2. Consider a source that produces an i.i.d. sequence of symbols from the alphabet \( \{A, B\} \) with probabilities 15/16, 1/16, respectively.
   (a) Find a Tunstall code for the source that encodes the source phrases into binary codewords of length 3.
   (b) Find a Tunstall code for the source that encodes the source phrases into binary codewords of length 4.

3. Using the Lempel Ziv window algorithm with a window size 32, encode the following text message.

```
MARY_HAD_A_LITTLE_LAMB_LITTLE_LAMB_LITTLE_LAMB_MARY_HAD_A_LITTLE_LAMB_WITH_FLEECE_AS_WHITE_AS_SNOW WHERE_IS_MARY _AND_HER_LITTLE_LAMB_HERE_IS_MARY_BUT_WHERE_IS_THE_LAMB
```

Use the following code for the length of the match: match length = 1 \( \rightarrow \) 0, match length = 2 \( \rightarrow \) 10, match length = 3 \( \rightarrow \) 110, match length = 4 \( \rightarrow \) 1110, etc.

Count the average number of encoded binary digits per source letter.

4. Using the algorithm derived in class, find the optimal quantization regions and representation points for the cases of 2, 3, 4 bit quantization for a random variable with probability density function

\[
f_X(x) = \frac{1}{2} e^{-|x|} \quad \text{for all } x.
\]

In each case, calculate the average mean squared error between the unquantized and quantized values.