CSC 310: Information Theory **Winter 2026**

Quiz 1: Entropy Calculation and Symbol Codes

Instructor: Swastik Kopparty *TA: Ziyang Jin*

Date: 14 Jan 2026

1.1 Questions (10 Points)

Let X be a discrete random variable with alphabet $\Sigma = \{a, b, c\}$, where:

$$X = \begin{cases} a & \text{w.prob. } \frac{2}{3}, \\ b & \text{w.prob. } \frac{1}{6}, \\ c & \text{w.prob. } \frac{1}{6}. \end{cases}$$

1. (2 points) What is the entropy $H(X)$? (Keep the log's in your answer; do not convert to numerical value).

   Solution.
   $$H(X) = \frac{2}{3} \log \frac{3}{2} + 2 \times \frac{1}{6} \log 6 = \frac{2}{3} \log \frac{3}{2} + \frac{1}{3} \log 6.$$

   If you try to simplify but did it wrong, you lose 0.5 points. A correct simplification would be $(\log 3) - \frac{1}{3}$, but why bother?

2. (3 points) Give a prefix-free symbol code $E : \Sigma \to \{0, 1\}^*$ for X which has expected value of $|E(X)|$ as small as you can make it. You do not need to prove that it is the smallest.

   Solution. $E(a) = \texttt{0},\ E(b) = \texttt{10},\ E(c) = \texttt{11}$ (or you switch $E(b)$ and $E(c)$).

   The other solution is $E(a) = \texttt{1},\ E(b) = \texttt{00},\ E(c) = \texttt{01}$ (or you switch $E(b)$ and $E(c)$).

   We accept all correct solutions.

3. (2 points) Suppose we had n independently sampled copies of X, denoted by $X_1 \ldots X_n$. What can you say about the (a) typical, (b) worst-case length of the concatenated string $E(X_1) \ldots E(X_n)$?

   Solution. Typical length is
   $$n \cdot \left(\frac{2}{3} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 2 \right) = \frac{4}{3} n.$$

   Some people write $n \cdot H(X)$ here, which is incorrect (-1 point). Because we ask the length of the symbol code. This one does not match the entropy.

   Worst-case length is $2n$ (like *bbbbbbb*... or *cccccc*.., etc.)

4. (3 points) Give a uniquely decodable symbol code C for X that is not prefix-free. Note that it need not minimize expected value of $|C(X)|$.

   Solution.

   There are many solutions. For example, $C(a) = \texttt{0},\ C(b) = \texttt{01},\ C(c) = \texttt{11}$. Observe this is not prefix-free because $C(a) = \texttt{0}$ is a prefix of $C(b) = \texttt{01}$. However, this code is suffix-free; no codeword is the suffix of another codeword.

   We accept all alternative solutions. For example, $C(a) = \texttt{1},\ C(b) = \texttt{100},\ C(c) = \texttt{101}$.