CSC 310: Information Theory **Winter 2026**

Tutorial 4: Simulate Sampling from a Given Distribution

Instructor: Swastik Kopparty *TA: Ziyang Jin*

Date: 28 Jan 2026

4.1 Review

We review some definitions from last tutorial.

Definition 4.1 (Joint Entropy) *The joint entropy $H(X, Y)$ of a pair of discrete random variables (X, Y) with a joint distribution $p(x, y)$ is defined as*

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log(p(x, y))$$

which can also be expressed as $H(X, Y) = -\mathbb{E}[\log(p(X, Y)]$.

Definition 4.2 (Conditional Entropy) *If random variables $(X, Y) \sim p(x, y)$, the conditional entropy $H(X|Y)$ is defined as*

$$H(X|Y) = \mathbb{E}_{y \in \mathcal{Y}}[H(X|Y = y)] = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y)$$

where $p(y) = \Pr[Y = y]$ and $\mathcal{Y}$ is the range of Y.

Theorem 4.3 *Let (X, Y) be a pair of discrete random variables, then the following is true*

1. *$H(X, Y) \leq H(X) + H(Y)$ with equality iff X, Y are independent.*
2. *$H(X, Y) = H(Y) + H(X|Y)$. (Chain Rule)*
3. *$H(X|Y) \leq H(X)$.*

4.2 Quiz Time

We will take a 15-minute quiz.

4.3 Simulate Sampling from a Given Distribution

Suppose you are given a supply of independent uniformly random bits. How would you simulate sampling from a distribution whose probability masses are $p_1, p_2, \ldots, p_n$ (which are real numbers)?

Answer. We use the random bits $b_1, b_2, \ldots$ to create a real number $x = 0.b_1b_2b_3\ldots$ (written in base 2). Then we see where x lies in the unit interval $[0, 1]$. We partition the unit interval into intervals of length $p_1, p_2, \ldots, p_n$, e.g., intervals $[0, p_1), [p_1, p_1 + p_2), [p_1 + p_2, p_1 + p_2 + p_3), \ldots, [1 - p_n - p_{n-1}, 1 - p_n), [1 - p_n, 1]$ and depending on which interval x lies in, we decide which outcome to take.

How many random bits do you have to see before you know which outcome to take?

Answer. It depends.

- Consider the example where $p_1, p_2, p_3 = \frac{1}{2}, \frac{1}{4}, \frac{1}{4}$. For this one, we only need 2 bits.

  Note that the values of x can be: (binary) `0.00` which is 0; (binary) `0.01` which is 0.25; (binary) `0.10` which is 0.5; (binary) `0.11` which is 0.75.

  The intervals are $I_1 = [0, 0.5), I_2 = [0.5, 0.75), I_3 = [0.75, 1]$. There is $\frac{1}{2}$ chance that $x \in I_1$, and $\frac{1}{4}$ chance that $x \in I_2$, and $\frac{1}{4}$ chance that $x \in I_3$.

- Consider another example where $p_1, p_2, p_3 = \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$. The intervals are $I_1 = [0, \frac{1}{3}), I_2 = [\frac{1}{3}, \frac{2}{3}), I_3 = [\frac{2}{3}, 1]$. For this one, we need infinitely many bits to simulate it since the power of 2 is never a multiple of 3.

  If we use 3 bits, x can take values like $0, \frac{1}{8}, \frac{2}{8}, \ldots, \frac{7}{8}$. There are 3 values $0, \frac{1}{8}, \frac{2}{8}$ lie in I_1, 3 values $\frac{3}{8}, \frac{4}{8}, \frac{5}{8}$ lie in I_2, and 2 values $\frac{6}{8}, \frac{7}{8}$ lie in I_3. The simulated distribution is actually $p_1 = \frac{3}{8}, p_2 = \frac{3}{8}, p_3 = \frac{1}{4}$. The error is $\frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

  If we use n bits, then there are 2^n values of x. With $\lceil \frac{2^n}{3} \rceil$ values lie in I_1, $\lceil \frac{2^n}{3} \rceil$ if n is even ($\lfloor \frac{2^n}{3} \rfloor$ if n is odd) values lie in I_2, and $\lfloor \frac{2^n}{3} \rfloor$ values lie in I_3.