EE5143 Problem Set 1: Entropy and Source Coding

Instructor: Dr. Andrew Thangaraj

Jan - May 2024

Entropy

1. *Entropy of functions.* (EIT 2.2) Let X be a random variable taking on a finite number of values. What is the (general) inequality relationship between $H(X)$ and $H(Y)$ if
   (a) $Y = 2^X$?
   (b) $Y = \cos X$?

2. *Entropy of a disjoint mixture.* (EIT 2.10) Let X_1 and X_2 be discrete random variables drawn according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over the respective alphabets $X_1 = \{1, 2, ..., m\}$ and $X_2 = \{m + 1, ..., n\}$. Let

$$X = \begin{cases} X_1 & \text{w.p. } \alpha \\ X_2 & \text{w.p. } 1 - \alpha \end{cases}$$

   (a) Find $H(X)$ in terms of $H(X_1)$, $H(X_2)$, and α.
   (b) Maximize over α to show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

3. *Entropy of a sum.* (EIT 2.14) Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectievely. Let $Z = X + Y$.
   (a) Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus, the addition of *independent* random variables adds uncertainty.
   (b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.
   (c) Under what conditions does $H(Z) = H(X) + H(Y)$?

4. *Mixing increases entropy.* Refer to EIT 2.28.

5. *Infinite Entropy.* (EIT 2.19) This problem shows that the entropy of a discrete random variable can be infinite. Let $A = \sum_{n=2}^{\infty}(n \log^2 n)^{-1}$. It is easy to show that A is finite by bounding the infinite sum by the integral of $(x \log^2 x)^{-1}$.

   Now, show that the integer-values random variable X defined by $Pr(X = n) = (An \log^2 n)^{-1}$ for $n = 2, 3, \ldots$ has $H(X) = +\infty$.

6. *Joint Entropy.* Let X and Y be random variables with joint pmf $p(x, y)$ as shown in the table below. Find $H(X)$, $H(Y)$, $H(Y|X)$, and $H(X, Y)$.

$X;Y$	0	1	2	3
0	1/16	1/8	1/32	1/32
1	1/8	1/16	1/32	1/32
2	1/16	1/16	1/16	1/16
3	1/4	0	0	0

7. *The value of a question.* (EIT 2.38) Let $X \sim p(x), x = 1, 2, \ldots, m$. We are given a set $S \subseteq \{1, 2. \ldots, m\}$. We ask whether $X \in S$ and receive the answer

$$Y = \begin{cases} 1 & \text{if } X \in S, \\ 0 & \text{if } X \notin S. \end{cases}$$

Suppose that $Pr\{X \in S\} = \alpha$. Find the decrease in uncertainty $H(X) - H(X|Y)$. Apparently, any set S with a given α is as good as any other.

8. *Entropy inequalities.* For random variables X, Y, Z, show the following:

   (a) $H(X,Y) + H(X,Y,Z) \leq H(X) + 2H(Y) + H(X,Z)$

   (b) $H(X,Y) + H(Y,Z) + H(X,Z) \geq 2H(X,Y,Z)$

9. *Coin Flips.* Refer to EIT 2.1

Optional Practice Problems.

2.4, 2.5, 2.6, 2.7, 2.8, 2.11, 2.13, 2.16, 2.18, 2.33, 2.35, 2.36, 2.41, 2.47 from EIT text book.

Source Coding

1. *Huffman Coding.*

   (a) (EIT 5.5) Find the binary Huffman code for the source with probabilities $\left(\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15}\right)$. Argue that this code is also optimal for the source with probabilities $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$.

   (b) (EIT 5.15) Construct a binary Huffman code for the following distribution on five symbols: $\mathbf{p} = (0.3, 0.3, 0.2, 0.1, 0.1)$. What is the average length of this code? Next, construct a probability distribution $\mathbf{p}'$ on five symbols for which the code that you constructed previously has an average length (under $\mathbf{p}'$) equal to its entropy $H(\mathbf{p}')$.

   (c) (EIT 5.14) Find the ternary Huffman codes for the random variable X with probabilities $p = \left(\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}\right)$ and calculate $L = \sum p_i l_i$.

   (d) (EIT 5.44*) Find the word lengths of the optimal binary encoding of $p = \left(\frac{1}{100}, \frac{1}{100}, \cdots \frac{1}{100}\right)$.

   (e) Now, consider a black box containing one of two possible sources A or B. Source A emits symbols with the distribution $P_A = \{\overset{1/4}{0}, \overset{1/4}{1}, \overset{1/4}{2}, \overset{1/4}{3}\}$, while Source B has distribution $P_B = \{\overset{1/2}{0}, \overset{1/4}{1}, \overset{1/8}{2}, \overset{1/8}{3}\}$.

      i. Design binary Huffman codes C_A and C_B for Sources A and B, respectively. What are the average lengths?

      ii. A prefix binary source code C is designed to encode the symbols $\{0, 1, 2, 3\}$ coming out of the black box when the source inside it is unknown.

         A. Suppose C is chosen to be C_A when Source B is actually in the black box. What is the average length? Denote this as $L_{P_B}(C_A)$.

         B. Suppose C is chosen to be C_B when Source A is actually in the black box. What is the average length? Denote this as $L_{P_A}(C_B)$.

         C. Can you construct a binary, prefix code C such that $L(C) = \frac{1}{2}L_{P_A}(C) + \frac{1}{2}L_{P_B}(C)$ is minimized? What is the minimum $L(C)$?

2. *Codes.* (EIT 5.37) For the codes C_1, C_2, C_3, C_4 shown below, indicate which of the following are,

$$C_1 = \{00, 01, 0\}$$ a

$$C_2 = \{00, 01, 100, 101, 11\}$$ a, b

$$C_3 = \{0, 10, 110, 1110, \ldots\}$$ a, b

$$C_4 = \{0, 00, 000, 0000\}$$

   (a) Uniquely decodable?

   (b) Instantaneous?

3. *Uniquely decodable codes.* Determine which of the following codes are uniquely decodable:

   (a) $\{0, 10, 11\}$

   (b) $\{0, 01, 10\}$ ×

   (c) $\{00, 01, 10, 11\}$

   (d) $\{110, 10, 11, 00, 100\}$

   Pref?

4. *Huffman vs. Shannon.* F (EIT 5.33) A random variable X takes on three values with probabilities $0.6, 0.3$ and 0.1.

   (a) What are the lengths of the binary Huffman codewords for X? What are the lengths of the binary Shannon codewords $\left(l(x) = \lceil \log \frac{1}{p(x)} \rceil\right)$ for X?

   (b) What is the smallest integer D such that the expected Shannon codeword length with a $D-$ary alphabet equals the expected Huffman codeword length with a $D-$ary alphabet?

5. *Bad wine.* (EIT 5.32) One is given 6 bottles of wine. It is known that precisely one bottle has gone bad (tastes terrible). From inspection of the bottles it is determined that the probability p_i that the i^{th} bottle is bad is given by $(p_1, p_2, \ldots p_6) = (\frac{7}{26}, \frac{6}{26}, \frac{4}{26}, \frac{4}{26}, \frac{3}{26}, \frac{3}{26})$. Tasting will determine the bad wine.

   Suppose you taste the wines one at a time. Choose the order of tasting to minimize the expected number of tastings required to determine the bad bottle. Remember, if the first 5 wines pass the test you don't have to taste the last.

   (a) What is the expected number of tastings required?

   (b) Which bottle should be tasted first?

   Now you get smart. For the first sample, you mix some of the wines in a fresh glass and sample the mixture. You proceed, mixing and tasting, stopping when the bad bottle has been determined.

   (a) What is the minimum expected number of tastings required to determine the bad wine?

   (b) What mixture should be tasted first?

6. *Lossless compression.* Consider a sequence $\{X_1, X_2, \dots X_N\}$ of i.i.d. Bernoulli random variables where $X_i \sim Ber(0.5)$. This sequence of random variables can be compressed and represented as a sequence of 'repeats' or 'run-lengths'. For example, 0011100000... has the repeats '00', '111' and '00000'. Let us assume that the sequence has 5 repeats and denote the k-th repeat length by R_k. It follows that $N = \sum_{k=1}^{5} R_k$.

   In particular, the sequence $\{X_1, X_2, \dots X_N\}$ can be compressed as $\{X_1, R_1, R_2, \dots, R_5\}$. Here are some examples,

   - $(100100\dots) \mapsto (1, 1, 2, 1, 2, \dots)$
   - $(111001011\dots) \mapsto (1, 3, 2, 1, 1, 2, \dots)$
   - $(0011110111\dots) \mapsto (0, 2, 4, 1, 3, \dots)$

   (a) Calculate the entropy of the first repeat R_1.

   (b) Calculate the joint entropy of $(X_1, R_1, R_2, \dots, R_5)$.

   (c) Design an optimal prefix code for R_1 and calculate the expected code length.

7. *The Games.*

   (a) *Lower bound on number of questions.* Arjun plays a game with Aryan. Arjun chooses some tool from a Mechanical tool kit, and Aryan attempts to identify the tool Arjun picked with a series of yes-no questions. Suppose Aryan is clever enough to use the code achieving the minimal expected length with respect to Arjun's distribution. We observe that Aryan requires an average of 9 questions to determine the tool Arjun has picked. Find a rough lower bound to the number of tools in the tool kit.

   (b) *The game of Hi-Lo* (EIT 5.19)

      i. A computer generates a random number N according to a known pmf $p(n), n \in \{1, 2, \dots, 100\}$. Seetha asks a yes-no question, "Is $N = n$?" and is told *yes, you are too high (or you are too low).* She continues for a total of six questions. If she is right (i.e., she receives the answer *yes*) during the six question sequence, she receives a prize of value $v(N)$. How should she proceed to maximize her expected winnings?

      ii. Consider the following variation: $N \sim p(n)$, prize $= v(n)$, $p(n)$ known, as before. But arbitrary yes-no questions are asked sequentially until N is determined. Questions cost 1 unit each. How should Seetha proceed? What is the expected payoff? Note that "Determined" doesn't mean that a *yes* answer is received.

      iii. Continuing the previous part, what if $v(n)$ is fixed but $p(n)$ can be chosen by the computer (and then announced to Seetha)? The computer wishes to minimize the Seetha's expected return. What should $p(n)$ be? What is the expected return to the Seetha?

Optional Practice Problems. 5.6, 5.8, 5.10, 5.17, 5.27, 5.40 From EIT text book.

Optional Challenging Problems:

1. *Axiomatic definition of entropy.* (EIT 2.46) A function $f(x_1, x_2, ..., x_m)$ is said to be symmetric if it produces the same value irrespective of the order of its arguments. Now, let us assume a sequence of symmetric functions $H_m(p_1, p_2, ..., p_m)$ satisfies the following properties:

- Normalization: $H_2(\frac{1}{2}, \frac{1}{2}) = 1$,
- Continuity: $H_2(p, 1-p)$ is a continuous function of p,
- Grouping: $H_m(p_1, p_2, ..., p_m) = H_{m-1}(p_1 + p_2, p_3, ..., p_m) + (p_1 + p_2)H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$
  Then show that H_m must be of the form:

$$H_m(p_1, p_2, ..., p_m) = -\sum_{i=1}^{m} p_i \log(p_i) \qquad m = 2, 3, ...$$

2. *Combinatorial meaning of entropy, Yury Polyanskiy.* Fix $n \geq 1$ and $0 \leq k \leq n$. Let $p = \frac{k}{n}$ and define $T_p \subset \{0,1\}^n$ to be the set of all binary sequences with p fraction of ones. Show that if $k \in [1, n-1]$ then

$$|T_p| = \sqrt{\frac{1}{np(1-p)}} \exp\{nh(p)\} C(n,k)$$

where $C(n,k)$ is bounded by two universal constants $C_0 \leq C(n,k) \leq C_1$, and $h(\cdot)$ is the binary entropy. Conclude that for all $0 \leq k \leq n$ we have

$$\log |T_p| = nh(p) + O(\log n).$$

Hint: Use Stirling's approximation:

$$\mathrm{e}^{\frac{1}{12n+1}} \leq \frac{n!}{\sqrt{2\pi n}(n/\mathrm{e})^n} \leq \mathrm{e}^{\frac{1}{12n}}, \quad n \geq 1$$