# EE276: Homework #1 Solutions

Due on Friday Jan 19, 5pm - Gradescope entry code: 2P885N

1. **Example of joint entropy.** Let $p(x, y)$ be given by

| $X$ \ $Y$ | 0 | 1 |
|---|---|---|
| 0 | $\frac{1}{3}$ | $\frac{1}{3}$ |
| 1 | 0 | $\frac{1}{3}$ |

Find

(a) $H(X), H(Y)$.
(b) $H(X \mid Y), H(Y \mid X)$.
(c) $H(X, Y)$.
(d) $H(Y) - H(Y \mid X)$.
(e) $I(X; Y)$.
(f) Draw a Venn diagram for the quantities in (a) through (e).

**Solution:** *Example of joint entropy*

(a) $H(X) = \frac{2}{3} \log \frac{3}{2} + \frac{1}{3} \log 3 = 0.918$ bits $= H(Y)$.
(b) $H(X|Y) = \frac{1}{3} H(X|Y = 0) + \frac{2}{3} H(X|Y = 1) = 0.667$ bits $= H(Y|X)$.
(c) $H(X, Y) = 3 \times \frac{1}{3} \log 3 = 1.585$ bits.
(d) $H(Y) - H(Y|X) = 0.251$ bits.
(e) $I(X; Y) = H(Y) - H(Y|X) = 0.251$ bits.
(f) See Figure 1.

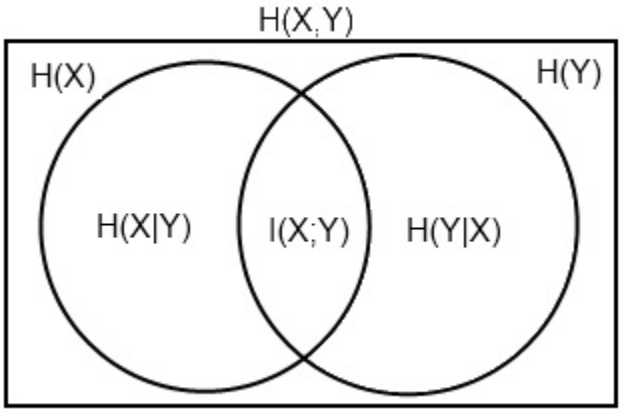
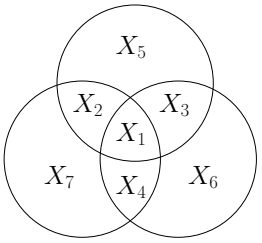
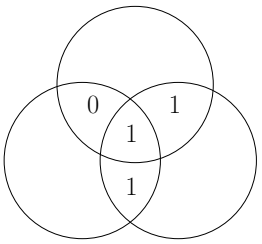



Figure 1: Venn diagram to illustrate the relationships of entropy and relative entropy

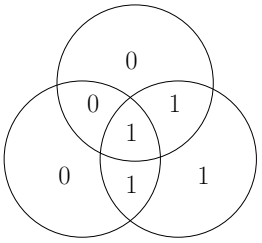
2. **Entropy of Hamming Code.**
   Hamming code is a simple error-correcting code that can correct up to one error in a sequence of bits. Now consider information bits $X_1, X_2, X_3, X_4 \in \{0, 1\}$ chosen uniformly at random, together with check bits $X_5, X_6, X_7$ chosen to make the parity of the circles even.
   (eg: $X_1 + X_2 + X_4 + X_7 = 0 \mod 2$)




   Thus, for example,




   becomes




That is, 1011 becomes 1011010.

(a) What is the entropy of $H(X_1, X_2, ..., X_7)$?

Now we make an error (or not) in one of the bits (or none). Let $\mathbf{Y} = \mathbf{X} \oplus \mathbf{e}$, where $\mathbf{e}$ is equally likely to be $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 0, 1)$, or $(0, 0, \ldots, 0)$, and $\mathbf{e}$ is independent of $\mathbf{X}$.

(b) Show that one can recover the message $\mathbf{X}$ perfectly from $\mathbf{Y}$. (Please provide a justification, detailed proof not required.)

(c) What is $H(\mathbf{X}|\mathbf{Y})$?

(d) What is $I(\mathbf{X}; \mathbf{Y})$?

(e) What is the entropy of $\mathbf{Y}$?

**Solution:** *Entropy of Hamming Code*

(a) By the chain rule,

$$H(X_1, X_2, ..., X_7) = H(X_1, X_2, X_3, X_4) + H(X_5, X_6, X_7|X_1, X_2, X_3, X_4).$$

Since $X_5, X_6, X_7$ are all deterministic functions of $X_1, X_2, X_3, X_4$, we have

$$H(X_5, X_6, X_7|X_1, X_2, X_3, X_4) = 0.$$

And since $X_1, X_2, X_3, X_4$ are independent Bernoulli(1/2) random variables,

$$H(X_1, X_2, ..., X_7) = H(X_1) + H(X_2) + H(X_3) + H(X_4) = 4.$$

(b) We first note that the Hamming code can detect one error. This follows from the fact that a flip of a single bit will result in the parity of at least one of the circles getting odd. Now, depending on which parities become odd, one can detect the precise location of the error. For example, if all three parities are odd, then $X_1$ is received in error. Similarly, if only the top circle parity is odd, $X_5$ is in error. In a similar manner, one can verify that $\mathbf{X}$ can be recoverd from $\mathbf{Y}$.

(c) As shown in (b), $\mathbf{X}$ is a deterministic function of $\mathbf{X} \oplus \mathbf{e}$. So $H(\mathbf{X}|\mathbf{Y}) = 0$.

(d) $I(\mathbf{X}; \mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) = 4$.

(e) We will expand $H(\mathbf{X} \oplus \mathbf{e}, \mathbf{X})$ in two different ways, using the chain rule. On one hand, we can write

$$\begin{aligned} H(\mathbf{X} \oplus \mathbf{e}, \mathbf{X}) &= H(\mathbf{X} \oplus \mathbf{e}) + H(\mathbf{X}|\mathbf{X} \oplus \mathbf{e}) \\ &= H(\mathbf{X} \oplus \mathbf{e}). \end{aligned}$$

In the last step, $H(\mathbf{X}|\mathbf{X} \oplus \mathbf{e}) = 0$ because $\mathbf{X}$ is a deterministic function of $\mathbf{X} \oplus \mathbf{e}$. On the other hand, we can also expand $H(\mathbf{X} \oplus \mathbf{e}, \mathbf{X})$ as follows:

$$\begin{aligned} H(\mathbf{X} \oplus \mathbf{e}, \mathbf{X}) &= H(\mathbf{X}) + H(\mathbf{X} \oplus \mathbf{e}|\mathbf{X}) \\ &= H(\mathbf{X}) + H(\mathbf{X} \oplus \mathbf{e} \oplus \mathbf{X}|\mathbf{X}) \\ &= H(\mathbf{X}) + H(\mathbf{e}|\mathbf{X}) \\ &= H(\mathbf{X}) + H(\mathbf{e}) \\ &= 4 + H(\mathbf{e}) \\ &= 4 + \log_2 8 \\ &= 7. \end{aligned}$$

The second equality follows since XORing with $\mathbf{X}$ is a one-to-one deterministic function (when conditioned on $\mathbf{X}$). The third equality follows from the well-known property of XOR that $y \oplus y = 0$. The fourth equality follows since the error vector $\mathbf{e}$ is independent of $\mathbf{X}$. The fifth equality follows since from part (a), we know that $H(\mathbf{X}) = 4$. The sixth equality follows since $\mathbf{e}$ is uniformly distributed over eight possible values: either there is an error in one of seven positions, or no error at all.

Equating our two different expansions for $H(\mathbf{X} \oplus \mathbf{e}, \mathbf{X})$, we have

$$H(\mathbf{X} \oplus \mathbf{e}, \mathbf{X}) = H(\mathbf{X} \oplus \mathbf{e}) = 7.$$

The entropy of $\mathbf{Y} = \mathbf{X} \oplus \mathbf{e}$ is 7 bits.

This result is closely related to the fact that the code in consideration, the Hamming [7,4,3] code, is a perfect code (`https://en.wikipedia.org/wiki/Hamming_bound#Perfect_codes`), and hence $\mathbf{X} \oplus \mathbf{e}$ is uniformly distributed in $\{0, 1\}^7$.

3. **Entropy of functions of a random variable.** Let $X$ be a discrete random variable. Show that the entropy of a function of $X$ is less than or equal to the entropy of $X$ by justifying the following steps:

$$\begin{aligned} H(X, g(X)) &\overset{(a)}{=} H(X) + H(g(X) \mid X) && (1) \\ &\overset{(b)}{=} H(X); && (2) \\ H(X, g(X)) &\overset{(c)}{=} H(g(X)) + H(X \mid g(X)) && (3) \\ &\overset{(d)}{\geq} H(g(X)). && (4) \end{aligned}$$

Thus $H(g(X)) \leq H(X)$.

**Solution:** *Entropy of functions of a random variable.*

(a) $H(X, g(X)) = H(X) + H(g(X)|X)$ by the chain rule for entropies.

(b) $H(g(X)|X) = 0$ since for any particular value of X, g(X) is fixed, and hence $H(g(X)|X) = \sum_x p(x) H(g(X)|X = x) = \sum_x 0 = 0$.

(c) $H(X, g(X)) = H(g(X)) + H(X|g(X))$ again by the chain rule.

(d) $H(X|g(X)) \geq 0$, with equality iff $X$ is a function of $g(X)$, i.e., $g(.)$ is one-to-one. Hence $H(X, g(X)) \geq H(g(X))$.

Combining parts (b) and (d), we obtain $H(X) \geq H(g(X))$.

4. **Coin flips.** A fair coin is flipped until the first head occurs. Let $X$ denote the number of flips required.

(a) Find the entropy $H(X)$ in bits. The following expressions may be useful:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \qquad \sum_{n=0}^{\infty} n r^n = \frac{r}{(1-r)^2}.$$

(b) A random variable $X$ is drawn according to this distribution. Construct an "efficient" sequence of yes-no questions of the form, "Is $X$ contained in the set $S$?" that determine the value of $X$. Compare $H(X)$ to the expected number of questions required to determine $X$.

**Solution:**

(a) The number $X$ of tosses till the first head appears has the geometric distribution with parameter $p = 1/2$, where $P(X = n) = pq^{n-1}$, $n \in \{1, 2, \ldots\}$. Hence the entropy of $X$ is

$$\begin{aligned} H(X) &= -\sum_{n=1}^{\infty} pq^{n-1} \log(pq^{n-1}) \\ &= -\left[\sum_{n=0}^{\infty} pq^n \log p + \sum_{n=0}^{\infty} npq^n \log q\right] \\ &= \frac{-p \log p}{1-q} - \frac{pq \log q}{p^2} \\ &= \frac{-p \log p - q \log q}{p} \\ &= H(p)/p \text{ bits.} \end{aligned}$$

If $p = 1/2$, then $H(X) = 2$ bits.

(b) Intuitively, it seems clear that the best questions are those that have equally likely chances of receiving a yes or a no answer. Consequently, one possible guess is that the most "efficient" series of questions is: Is $X = 1$? If not, is $X = 2$? If not, is $X = 3$? ...with a resulting expected number of questions equal to $\sum_{n=1}^{\infty} n(1/2^n) = 2$. This should reinforce the intuition that $H(X)$ is

a measure of the uncertainty of $X$. Indeed in this case, the entropy is exactly the same as the average number of questions needed to define $X$, and in general $E(\#\text{ of questions}) \geq H(X)$. This problem has an interpretation as a source coding problem. Let 0 =no, 1 =yes, $X$ =Source, and $Y$ =Encoded Source. Then the set of questions in the above procedure can be written as a collection of $(X, Y)$ pairs: $(1, 1)$, $(2, 01)$, $(3, 001)$, etc. . In fact, this intuitively derived code is the optimal (Huffman) code minimizing the expected number of questions.

5. **Minimum entropy.** In the following, we use $H(p_1, ..., p_n) \equiv H(\mathbf{p})$ to denote the entropy $H(X)$ of a random variable $X$ with alphabet $\mathcal{X} := \{1, \ldots, n\}$, i.e.,

$$H(X) = -\sum_{i=1}^{n} p_i \log(p_i).$$

What is the minimum value of $H(p_1, ..., p_n) = H(\mathbf{p})$ as $\mathbf{p}$ ranges over the set of $n$-dimensional probability vectors? Find all $\mathbf{p}$'s which achieve this minimum.

**Solution:** We wish to find *all* probability vectors $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ which minimize

$$H(\mathbf{p}) = -\sum_{i} p_i \log p_i.$$

Now $-p_i \log p_i \geq 0$, with equality iff $p_i = 0$ or 1. Hence the only possible probability vectors which minimize $H(\mathbf{p})$ are those with $p_i = 1$ for some $i$ and $p_j = 0, j \neq i$. There are $n$ such vectors, i.e., $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, $\ldots$, $(0, \ldots, 0, 1)$, and the minimum value of $H(\mathbf{p})$ is 0.

6. **Drawing with and without replacement.** An urn contains $r$ red, $w$ white, and $b$ black balls. Which has higher entropy, drawing $k \geq 2$ balls from the urn with replacement or without replacement? Set it up and show why. (There is both a hard way and a relatively simple way to do this.)

**Solution:** *Drawing with and without replacement.* Intuitively, it is clear that if the balls are drawn with replacement, the number of possible choices for the $i$-th ball is larger, and therefore the conditional entropy is larger. But computing the conditional distributions is slightly involved. It is easier to compute the unconditional entropy.

- With replacement. In this case the conditional distribution of each draw is the same for every draw. Thus

$$X_i = \begin{cases} \text{red} & \text{with prob.} \frac{r}{r+w+b} \\ \text{white} & \text{with prob.} \frac{w}{r+w+b} \\ \text{black} & \text{with prob.} \frac{b}{r+w+b} \end{cases} \tag{5}$$

and therefore

$$H(X_i|X_{i-1}, \ldots, X_1) = H(X_i) \tag{6}$$

$$= \log(r+w+b) - \frac{r}{r+w+b}\log r - \frac{w}{r+w+b}\log w - \frac{b}{r+w+b}\log b. \tag{7}$$

- Without replacement. The unconditional probability of the $i$-th ball being red is still $r/(r+w+b)$, etc. Thus the unconditional entropy $H(X_i)$ is still the same as with replacement. The conditional entropy $H(X_i|X_{i-1},\ldots,X_1)$ is less than the unconditional entropy, and therefore the entropy of drawing without replacement is lower.

7. **Infinite entropy.**
This problem shows that the entropy of a discrete random variable can be infinite. (In this question you can take log as the natural logarithm for simplicity.)

   (a) Let $A = \sum_{n=2}^{\infty}(n\log^2 n)^{-1}$. Show that $A$ is finite by bounding the infinite sum by the integral of $(x\log^2 x)^{-1}$.

   (b) Show that the integer-valued random variable $X$ distributed as:
   $P(X=n) = (An\log^2 n)^{-1}$ for $n = 2, 3, \ldots$ has entropy $H(X)$ given by:

$$H(X) = \log A + \sum_{n=2}^{\infty}\frac{1}{An\log n} + \sum_{n=2}^{\infty}\frac{2\log\log n}{An\log^2 n}$$

   (c) Show that the entropy $H(X) = +\infty$ (by showing that the sum $\sum_{n=2}^{\infty}\frac{1}{n\log n}$ diverges).

**Solution: Infinite entropy.**
We use the technique of bounding sums by integrals, see `https://math.stackexchange.com/questions/1282807/bounding-a-summation-by-an-integral` for an example with some figures.

(a) Define a function $f : [2,\infty) \to \mathbb{R}$ as follows:

$$f(x) = (\lceil x\rceil \log^2 \lceil x\rceil)^{-1}$$

Then, $f(x) \le (x\log^2 x)^{-1}$ and

$$\begin{aligned}
A &= (2\log^2 2)^{-1} + \sum_{n=3}^{\infty}(n\log^2 n)^{-1} \\
&= (2\log^2 2)^{-1} + \int_2^{\infty}(\lceil x\rceil \log^2 \lceil x\rceil)^{-1}dx \\
&\le (2\log^2 2)^{-1} + \int_2^{\infty}(x\log^2 x)^{-1}dx \\
&= (2\log^2 2)^{-1} + \frac{1}{\log 2} \\
&< \infty
\end{aligned}$$

(b) By definition, $p_n = \Pr(X = n) = 1/An\log^2 n$ for $n \geq 2$. Therefore

$$\begin{aligned} H(X) &= -\sum_{n=2}^{\infty} p_n \log p_n \\ &= -\sum_{n=2}^{\infty} \left(1/An\log^2 n\right) \log\left(1/An\log^2 n\right) \\ &= \sum_{n=2}^{\infty} \frac{\log(An\log^2 n)}{An\log^2 n} \\ &= \sum_{n=2}^{\infty} \frac{\log A + \log n + 2\log\log n}{An\log^2 n} \\ &= \log A + \sum_{n=2}^{\infty} \frac{1}{An\log n} + \sum_{n=2}^{\infty} \frac{2\log\log n}{An\log^2 n}\,. \end{aligned}$$

(c) The first term is finite. For base 2 logarithms, all the elements in the sum in the last term are nonnegative. (For any other base, the terms of the last sum eventually all become positive.) So all we have to do is bound the middle sum, which we do by comparing with an integral (in a similar manner as done in part (a), here using $\lfloor x \rfloor$ instead of $\lceil x \rceil$).

$$\sum_{n=2}^{\infty} \frac{1}{An\log n} = \int_2^{\infty} \frac{1}{A\lfloor x\rfloor \log\lfloor x\rfloor} dx > \int_2^{\infty} \frac{1}{Ax\log x} dx = K \ln\ln x \,\Big|_2^{\infty} = +\infty\,.$$

We conclude that $H(X) = +\infty$.