EE276: Homework #1 Solutions

Due on Friday Jan 16, 6pm - Gradescope entry code: E6VP4X

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

X \ Y	0	1
0	$\frac{1}{10}$	$\frac{3}{10}$
1	$\frac{2}{10}$	$\frac{4}{10}$

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 relating the quantities in (a) through (e).

Numerically round the answers to three decimal places.

Solution:

(a) $H(X) = \frac{4}{10}\log\frac{10}{4} + \frac{6}{10}\log\frac{10}{6} \approx 0.97$ bit.
$H(Y) = \frac{3}{10}\log\frac{10}{3} + \frac{7}{10}\log\frac{10}{7} \approx 0.88$ bits.

(b) $H(X|Y) = \frac{3}{10}H(X|Y=0) + \frac{7}{10}H(X|Y=1) = 0.3 \cdot 0.918 + 0.7 \cdot 0.985 \approx 0.965$ bits.
$H(Y|X) = \frac{4}{10}H(Y|X=0) + \frac{6}{10}H(Y|X=1) = 0.4 \cdot 0.811 + 0.6 \cdot 0.918 \approx 0.875$ bits.

(c) $H(X,Y) = \frac{1}{10}\log\frac{1}{10} + \frac{3}{10}\log\frac{3}{10} + \frac{2}{10}\log\frac{2}{10} + \frac{4}{10}\log\frac{4}{10} \approx 1.846$ bits.

(d) $H(Y) - H(Y|X) \approx 0.006$ bits.

(e) $I(X;Y) = H(Y) - H(Y|X) \approx 0.006$ bits.

(f) See Figure 1.

2. **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

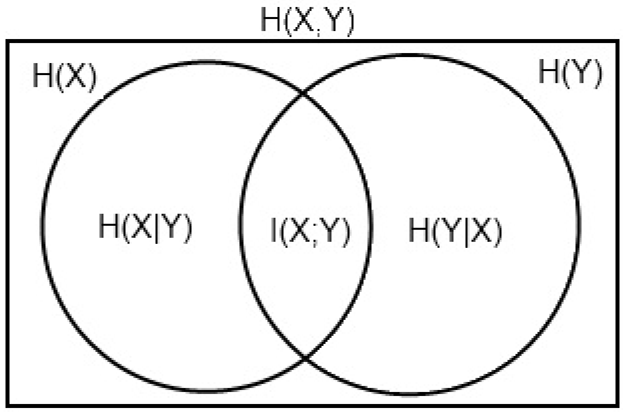

Figure 1: Venn diagram to illustrate the relationships of entropy and relative entropy

justifying the following steps:

$$H(X, g(X)) \stackrel{(a)}{=} H(X) + H(g(X) \mid X) \quad (1)$$
$$\stackrel{(b)}{=} H(X); \quad (2)$$
$$H(X, g(X)) \stackrel{(c)}{=} H(g(X)) + H(X \mid g(X)) \quad (3)$$
$$\stackrel{(d)}{\geq} H(g(X)). \quad (4)$$

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 $H(X) \geq H(g(X))$.

3. **Entropy of a disjoint mixture.** 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 $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m + 1, \ldots, n\}$. Let

$$X = \begin{cases} X_1, & \text{with probability } \alpha, \\ X_2, & \text{with probability } 1 - \alpha. \end{cases}$$

(a) Find $H(X)$ in terms of $H(X_1)$ and $H(X_2)$ 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.

Solution: *Entropy of a disjoint mixture.*

(a) We can do this problem by writing down the definition of entropy and expanding the various terms. Instead, we will use the algebra of entropies for a simpler proof. Since X_1 and X_2 have disjoint support sets, we can write

$$X = \begin{cases} X_1 & \text{with probability} \quad \alpha \\ X_2 & \text{with probability} \quad 1-\alpha \end{cases}$$

Define a function of X,

$$\theta = f(X) = \begin{cases} 1 & \text{when } X = X_1 \\ 2 & \text{when } X = X_2 \end{cases}$$

Then as in problem 1, we have

$$\begin{aligned} H(X) &= H(X, f(X)) = H(\theta) + H(X|\theta) \\ &= H(\theta) + p(\theta = 1)H(X|\theta = 1) + p(\theta = 2)H(X|\theta = 2) \\ &= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2) \end{aligned}$$

where $H(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$.

(b) We first find the maximum value of $H(X)$ in terms of $H(X_1)$ and $H(X_2)$ by maximizing over α. Taking the derivative of the expression for $H(X)$ from part (a) with respect to α gives

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(-\alpha \log \alpha - (1-\alpha)\log(1-\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2)\right) \\ = \; & -\log \alpha + \log(1-\alpha) + H(X_1) - H(X_2) \end{aligned}$$

Second derivative of $H(X)$ with respect to α becomes

$$-\frac{1}{\alpha} - \frac{1}{1-\alpha},$$

which is negative for all $\alpha \in (0,1)$, and $H(X)$ is continuous as a function of α. This implies a local minimum of the function is also a global minimum. Then, by setting its derivative to zero and solving for the corresponding value of α (hereafter called α^*), we can obtain the maximum value $H(X)$ takes as a function of α in terms of $H(X_1)$ and $H(X_2)$:

$$\begin{aligned} & -\log \alpha^* + \log(1-\alpha^*) + H(X_1) - H(X_2) = 0 \\ \implies \; & H(X_1) - \log \alpha^* = H(X_2) - \log(1-\alpha^*) \\ \implies \; & \alpha^* = \frac{2^{H(X_1)-H(X_2)}}{1 + 2^{H(X_1)-H(X_2)}}. \end{aligned}$$

Finally, using the equations above, for any α, we can write

$$\begin{aligned} H(X) &\le H(X)|_{\alpha=\alpha^*} \\ &= \alpha^*(H(X_1) - \log \alpha^*) + (1-\alpha^*)(H(X_2) - \log(1-\alpha^*)) \\ &= H(X_1) - \log \alpha^*, \end{aligned}$$

which implies

$$\begin{aligned} 2^{H(X)} &\le 2^{H(X_1)} \frac{1}{\alpha^*} \\ &= 2^{H(X_1)} \frac{1 + 2^{H(X_1)-H(X_2)}}{2^{H(X_1)-H(X_2)}} \\ &= 2^{H(X_1)} + 2^{H(X_2)}. \end{aligned}$$

Since $2^{H(X)}$ is the effective alphabet size for the random variable X, the inequality shows that the effective alphabet size of X is less than or equal to the sum of the effective alphabet sizes for X_1 and X_2, with equality satisfied at the value of $\alpha = \alpha^*$ that maximizes the entropy of 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., $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, ..., $(0, \ldots, 0, 1)$, and the minimum value of $H(\mathbf{p})$ is 0.

6. **Mixing increases entropy.** Let $p_i > 0, \quad i = 1, 2, \ldots m$. Show that the entropy of a random variable distributed according to $(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)$, is less than or equal to the entropy of a random variable distributed according to $(p_1, \ldots, \frac{p_i+p_j}{2}, \ldots, \frac{p_i+p_j}{2}, \ldots, p_m)$.

Solution: This problem depends on the convexity of the log function. Let

$$\begin{aligned} P_1 &= (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m) \\ P_2 &= (p_1, \ldots, \frac{p_i + p_j}{2}, \ldots, \frac{p_j + p_i}{2}, \ldots, p_m) \end{aligned}$$

Then,

$$\begin{aligned} H(P_2) - H(P_1) &= -2\left(\frac{p_i + p_j}{2}\right) \log\left(\frac{p_i + p_j}{2}\right) + p_i \log p_i + p_j \log p_j \\ &= -(p_i + p_j) \log\left(\frac{p_i + p_j}{2}\right) + p_i \log p_i + p_j \log p_j. \end{aligned}$$

Since $f(x) = x \log x$ is convex in $x > 0$, the following inequality holds:

$$\frac{f(p_i) + f(p_j)}{2} = \frac{1}{2}(p_i \log p_i + p_j \log p_j) \geq f\left(\frac{p_i + p_j}{2}\right) = \frac{p_i + p_j}{2} \log \frac{p_i + p_j}{2}.$$

Thus,

$$H(P_2) \geq H(P_1).$$

7. **Infinite entropy. [Bonus]**
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: 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, $f(x) \leq (x \log^2 x)^{-1}$ and

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

(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 $H(X) = +\infty$.