ASSIGNMENT 3

Exercise 1. (Mixing increases entropy) Show that the entropy of the probability distribution, $(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)$ is less than that of the distribution $\left(p_1, \ldots, \frac{p_i+p_j}{2}, \ldots, \frac{p_i+p_j}{2}, \ldots, p_m\right)$.

Solution. Let $P \equiv (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)$ and $Q \equiv \left(p_1, \ldots, \frac{p_i+p_j}{2}, \ldots, \frac{p_i+p_j}{2}, \ldots, p_m\right)$. Then,

$$\begin{aligned} H(Q) - H(P) &= 2\left(\frac{p_i + p_j}{2}\right) \log\left(\frac{2}{p_i + p_j}\right) - p_i \log \frac{1}{p_i} - p_j \log \frac{1}{p_j} \\ &= p_i \log \frac{2p_i}{(p_i + p_j)} + p_j \log \frac{2p_j}{p_i + p_j} \\ &= (p_i + p_j) \left[\frac{p_i}{p_i + p_j} \log \frac{p_i/(p_i + p_j)}{1/2} + \frac{p_j}{p_i + p_j} \log \frac{p_j/(p_i + p_j)}{1/2}\right]. \end{aligned}$$

Identify the expression on the right side as $(p_i + p_j)$ times the KL divergence between Bernoulli distributions $\left(\frac{p_i}{p_i+p_j}, \frac{p_j}{p_i+p_j}\right)$ and $\left(\frac{1}{2}, \frac{1}{2}\right)$, which is non-negative. □

Exercise 2. (Entropy of common distributions) Calculate the entropy of X where

a. X is the output of n independent tosses of a coin which shows heads with probability p.

b. X is a *Geo*(p) random variable. That is, $\mathbb{P}[X = k] = (1 - p)^{k-1}p$.

Solution. a. $H(X) = H(X_1, X_2, ..., X_n)$ where $X_i \sim$ i.i.d. $Ber(p)$. Therefore,

$$H(X) = nH(X_1) = n[-p \log p - (1 - p) \log(1 - p)].$$

b. $X \sim Geo(p)$. We know that for $X \sim Geo(p), \mathbb{E}[X] = \frac{1}{p}$. Let $h(X) = -\log P(X)$ where $P(k) = (1 - p)^{k-1}p$. Then,

$$\begin{aligned} H(X) &= \mathbb{E}[h(X)] \\ &= \mathbb{E}[-\log\{(1 - p)^{X-1}p\}] \\ &= \mathbb{E}[(1 - X) \log(1 - p) - \log p] \\ &= \left(1 - \frac{1}{p}\right) \log(1 - p) - \log p \\ &= \frac{-(1 - p) \log(1 - p) - p \log p}{p} \end{aligned}$$

□

Exercise 3. (KL divergence) Calculate the KL divergence (relative entropy) between P and Q where

a. $P \equiv Geo(p)$ and $Q \equiv Geo(q)$.

b. $P \equiv \mathcal{N}(\mu_1, \sigma^2)$ and $Q \equiv \mathcal{N}(\mu_2, \sigma^2)$

Solution. a. $P \equiv Geo(p), Q \equiv Geo(q)$.

$$\begin{aligned}D(P||Q) &= \mathbb{E}_P\left[\log \frac{(1-p)^{X-1}p}{(1-q)^{X-1}q}\right] \\ &= \mathbb{E}_P\left[(X-1)\log\left(\frac{1-p}{1-q}\right) + \log\left(\frac{p}{q}\right)\right] \\ &= \left(\frac{1}{p} - 1\right)\log\left(\frac{1-p}{1-q}\right) + \log\left(\frac{p}{q}\right)\end{aligned}$$

a. $P \equiv \mathcal{N}(\mu_1, \sigma^2), Q \equiv \mathcal{N}(\mu_2, \sigma^2)$.

$$\begin{aligned}D(P||Q) &= \int_{\mathbb{R}} \frac{e^{\frac{-(x-\mu_1)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\left[\frac{(x-\mu_1)^2 - (x-\mu_2)^2}{2\sigma^2}\right]dx \\ &= \int_{\mathbb{R}} \frac{e^{\frac{-(x-\mu_1)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\left[\frac{2x(\mu_2-\mu_1) + \mu_2^2 - \mu_1^2}{2\sigma^2}\right]dx \\ &= \frac{2(\mu_2-\mu_1)}{2\sigma^2}\int_{\mathbb{R}} \frac{e^{\frac{-(x-\mu_1)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \cdot x \cdot dx + \frac{(\mu_2^2-\mu_1^2)}{2\sigma^2}\int_{\mathbb{R}} \frac{e^{\frac{-(x-\mu_1)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \cdot dx \\ &= \frac{2(\mu_2-\mu_1)\mu_1 + \mu_2^2 - \mu_1^2}{2\sigma^2} \\ &= \frac{(\mu_1-\mu_2)^2}{2\sigma^2}.\end{aligned}$$

□

Exercise 4 (Mutual information)**.** a. Let X be a uniform random variable over $\{1, 2, 3, 4\}$. Let

$$Y = \begin{cases} 0 \text{ if } X \text{ is odd} \\ 1 \text{ otherwise.} \end{cases} \qquad Z = \begin{cases} 0 \text{ if } X \text{ is even} \\ 1 \text{ otherwise.} \end{cases}$$

Find $I(Y; Z)$.

b. We roll a fair die which has six sides (opposite sides of a die add up to 7). What is the mutual information between the top side and the one facing you?

Solution. a. Note that always $Y \neq Z$, which means knowing Z lets us know Y, *i.e.* $H(Y|Z) = 0$.

$$I(Y; Z) = H(Y) - H(Y|Z) = 1 - 0 = 1.$$

b. Top side X_T can take any of $\{1, 2, 3, 4, 5, 6\}$ with same probability. Moreover, knowing the one facing us, X_F, X_T can take four values with same probability, so

$$I(X_T; X_F) = H(X_T) - H(X_T|X_F) = \log(6) - \log(4).$$

□

Exercise 5 (Entropy and Mutual Information)**.** Prove the following inequalities:

a. $H(X, Y|Z) \geq H(X|Z)$,

b. $I(X,Y;Z) \geq I(X;Z)$,

c. $H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X)$.

Solution. a.

$$\begin{aligned} H(X,Y|Z) &\overset{(a)}{=} H(X|Z) + H(Y|X,Z) \\ &\overset{(b)}{\geq} H(X|Z) \end{aligned}$$

where (a) holds by the chain rule for entropy and where (b) follows by the non-negativity of entropy.

b.

$$\begin{aligned} I(X,Y|Z) &\overset{(a)}{=} I(X;Z) + I(Y;Z|X) \\ &\overset{(b)}{\geq} I(X;Z) \end{aligned}$$

where (a) holds by the chain rule for mutual information and where (b) holds by the non-negativity of mutual information.

c.

$$\begin{aligned} H(X,Y,Z) - H(X,Y) &\overset{(a)}{=} (H(X,Z) + H(Y|X,Z)) - (H(X) + H(Y|X)) \\ &\overset{(b)}{\leq} H(X,Z) - H(X) \end{aligned}$$

where (a) is due to the chain rule for entropy and where (b) holds since conditioning cannot increase entropy.

□

Exercise 6 (Conditioning for mutual information)**.** Give examples of joint random variables X, Y, and Z such that

a. $I(X;Y|Z) < I(X;Y)$.

b. $I(X;Y|Z) > I(X;Y)$.

Solution. a. Let X be Bernoulli$\left(\frac{1}{2}\right)$ random variable and $Z = Y = X$. Then,

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|X) - H(X|X) = 0 - 0 = 0$$

$$I(X;Y) = H(X) - H(X|Y) = H(X) - H(X|X) = H(X) - 0 = 1.$$

b. Let X and Y be independent Bernoulli$\left(\frac{1}{2}\right)$ random variables and $Z = X + Y$. Then,

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X) - H(X|X,Y) = 1 - 0 = 1$$

$$I(X;Y) = H(X) - H(X|Y) = H(X) - H(X) = 0.$$

□

Exercise 7 (Entropy and pairwise independence)**.** Let X, Y, Z be three binary Bernoulli($\frac{1}{2}$) random variables that are pairwise independent; that is, $I(X;Y) = I(X;Z) = I(Y;Z) = 0$.

a. Under this constraint, what is the minimum value for $H(X,Y,Z)$?

b. Give an example achieving this minimum.

Solution. a.

$$\begin{aligned} H(X,Y,Z) &= H(X) + H(Y|X) + H(Z|Y,X) \\ &= H(X) + H(Y) + H(Z|Y,X) \\ &\geq H(X) + H(Y) \\ &= 2 \end{aligned}$$

b. Let $Z = X \oplus Y$. Verify that $I(X;Z) = I(Y;Z) = 0$.

□

Exercise 8. (Conditioning and sub additivity) Prove the following.

a.
$$H(X_1, X_2, X_3) \leq \frac{1}{2}\left[H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1)\right].$$

b.
$$H(X_1, X_2, X_3) \geq \frac{1}{2}\left[H(X_1, X_2|X_3) + H(X_2, X_3|X_1) + H(X_3, X_1|X_2)\right].$$

Solution. a. Using chain rule, $H(X_1, X_2, X_3)$ can be expanded in the following two ways.

$$\begin{aligned} 2H(X_1, X_2, X_3) &= H(X_1, X_2) + H(X_3|X_1, X_2) + H(X_2, X_3) + H(X_1|X_2, X_3) \\ &= H(X_1, X_2) + H(X_2, X_3) + H(X_3|X_1, X_2) + H(X_1|X_2, X_3) \\ &\leq H(X_1, X_2) + H(X_2, X_3) + H(X_3|X_1, X_2) + H(X_1) \\ &\leq H(X_1, X_2) + H(X_2, X_3) + H(X_3|X_1) + H(X_1) \\ &= H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1). \end{aligned}$$

b. Add and subtract $H(X_1) + H(X_2) + H(X_3)$.

$$\begin{aligned} &H(X_1, X_2|X_3) + H(X_2, X_3|X_1) + H(X_3, X_1|X_2) \\ &= H(X_1, X_2|X_3) + H(X_3) + H(X_2, X_3|X_1) + H(X_1) + H(X_3, X_1|X_2) + H(X_2) \\ &\quad - \big[H(X_1) + H(X_2) + H(X_3)\big] \\ &= 3H(X_1, X_2, X_3) - \big[H(X_1) + H(X_2) + H(X_3)\big] \\ &\leq 3H(X_1, X_2, X_3) - H(X_1, X_2, X_3) \\ &= 2H(X_1, X_2, X_3). \end{aligned}$$

□

Exercise 9. Show that among all $\mathbb{N}$-valued random variables X with $\mathbb{E}[X] = \mu$, the $Geo(1/\mu)$ random variable has the maximum value of Shannon entropy.
Hint – Consider random variables X and Y with mean μ and taking values in $\mathbb{N}$ with $X \sim P_X$ and $Y \sim P_Y$ where P_Y is Geometric, and calculate $D(P_X||P_Y)$.

Solution. Let X be a r.v. such that $X = i$ with probability $P_X(i), i \in \mathbb{N}$ and $\mathbb{E}[X] = \mu$. Let $Y \sim P_Y \equiv Geo\left(\frac{1}{\mu}\right)$. Therefore, $\mathbb{E}[Y] = \mu$. Then,

$$\begin{aligned} D(P_X||P_Y) &= \sum_{i=1}^{\infty} P_X(i) \log \frac{P_X(i)}{P_Y(i)} \\ &= \sum_{i=1}^{\infty} P_X(i) \log P_X(i) - P_Y(i) \log P_Y(i) + P_Y(i) \log P_Y(i) - P_X(i) \log P_Y(i) \\ &= H(Y) - H(X) + \sum_{i=1}^{\infty} \Big[P_Y(i) \log P_Y(i) - P_X(i) \log P_Y(i) \Big]. \end{aligned} \tag{1}$$

Since $P_Y(i) = \left(1 - \frac{1}{\mu}\right)^{i-1} \left(\frac{1}{\mu}\right)$,

$$\begin{aligned} \sum_{i=1}^{\infty} P_X(i) \log P_Y(i) &= \sum_{i=1}^{\infty} P_X(i) \cdot (i-1) \log(\mu - 1) - \sum_{i=1}^{\infty} P_X(i) \cdot i \cdot \log \mu \\ &= (\mu - 1) \log(\mu - 1) - \mu \log \mu. \end{aligned} \tag{2}$$

From the entropy calculation of a Geometric r.v. (Exer. $2b$), we know that

$$\begin{aligned} \sum_{i=1}^{\infty} P_Y(i) \log P_Y(i) &= \frac{\left(1 - \frac{1}{\mu}\right) \log \left(1 - \frac{1}{\mu}\right) + \left(\frac{1}{\mu}\right) \log \left(\frac{1}{\mu}\right)}{1/\mu} \\ &= (\mu - 1) \log(\mu - 1) - \mu \log \mu. \end{aligned} \tag{3}$$

Substituting (2) and (3) in (1), we get

$$\begin{aligned} H(Y) - H(X) &= D(P_X||P_Y) \\ &\geq 0. \end{aligned}$$

Therefore, for any r.v. $X \in \mathbb{N}$ with $\mathbb{E}[X] = \mu$, $H(X) \leq H(Y)$ where $Y \sim Geo\left(\frac{1}{\mu}\right)$.

□

Exercise 10 (Conditional mutual information)**.** Consider a sequence of n binary random variables $X_1, X_2, \cdots, X_n$. Each sequence with an even number of 1's has probability $2^{-(n-1)}$, and each sequence with an odd number of 1's has probability 0. Find the mutual informations $I(X_1; X_2)$, $I(X_2; X_3|X_1), \ldots, I(X_{n-1}; X_n|X_1, \ldots, X_{n-2})$.

Proof. We always have $X_n = X_1 \oplus X_2 \oplus \cdots \oplus X_{n-1}$,[1] since the sequences with odd number of ones have zero probability, and since each sequence with even number of 1s is equiprobable, X_1, X_2, $\cdots$, X_n are independent Bernoulli$\left(\frac{1}{2}\right)$ random variables. So, for $2 \leq i \leq n-2$,

$$\begin{aligned} I(X_i; X_{i+1}|X_1, \cdots, X_{i-1}) &= H(X_{i+1}|X_1, \cdots, X_{i-1}) - H(X_{i+1}|X_1, \cdots, X_{i-1}, X_i) \\ &= H(X_{i+1}) - H(X_{i+1}) = 0 \end{aligned}$$

[1] $\oplus$ is sum modulo 2.

and for $i = n-1$,

$$
\begin{aligned}
I(X_{n-1};X_n|X_1,\cdots,X_{n-2}) &= H(X_n|X_1,\cdots,X_{n-2}) - H(X_n|X_1,\cdots,X_{n-2},X_{n-1}) \\
&= H(X_1 \oplus X_2 \oplus \cdots \oplus X_{n-1}|X_1,\cdots,X_{n-2}) - H(X_1 \oplus X_2 \oplus \cdots \oplus X_{n-1}|X_1,\cdots,X_{n-2},X_{n-1}) \\
&= H(X_{n-1}|X_1,\cdots,X_{n-2}) - 0 \\
&= H(X_{n-1}) = 1
\end{aligned}
$$

□