Homework 2 -Solutions

Problem 1 Given $p_{X,Y}(0,0)=\frac{1}{4}, p_{X,Y}(0,1)=\frac{1}{4}, p_{X,Y}(1,0)=0, p_{X,Y}(1,1)=\frac{1}{2}$.
Preliminaries:

$$p_X(x=0)=p_{X,Y}(x=0,y=0)+p_{X,Y}(x=0,y=1)=\frac{1}{2}, p_X(x=1)=1-p_X(x=1)=\frac{1}{2}$$

$$p_Y(y=0)=p_{X,Y}(x=0,y=0)+p_{X,Y}(x=1,y=0)=\frac{1}{4}, p_Y(y=1)=1-p_Y(y=1)=\frac{3}{4}$$

$$p_{X|Y}(X=0|Y=0)=\frac{p_{X,Y}(X=0,Y=0)}{p_Y(Y=0)}=1, p_{X|Y}(X=1|Y=0)=1-p_{X|Y}(X=1|Y=0)=0$$

$$p_{X|Y}(X=0|Y=1)=\frac{p_{X,Y}(X=0,Y=1)}{p_Y(Y=1)}=\frac{1}{3}, p_{X|Y}(X=1|Y=1)=1-p_{X|Y}(X=1|Y=1)=\frac{2}{3}$$

$$p_{Y|X}(Y=0|X=0)=\frac{p_{X,Y}(X=0,Y=0)}{p_X(X=0)}=\frac{1}{2}, p_{Y|X}(Y=1|X=0)=1-p_{Y|X}(Y=1|X=0)=\frac{1}{2}$$

$$p_{Y|X}(Y=0|X=1)=\frac{p_{X,Y}(X=1,Y=0)}{p_X(X=1)}=0, p_{Y|X}(Y=1|X=0)=1-p_{Y|X}(Y=1|X=1)=$$

(a)

$$\begin{aligned}H(X)&=p_X(x=0)\log\frac{1}{p_X(x=0)}+p_X(x=1)\log\frac{1}{p_X(x=1)}\\&=\frac{1}{2}\log 2+\frac{1}{2}\log 2=\log 2=1\end{aligned}$$

$$\begin{aligned}H(Y)&=p_Y(y=0)\log\frac{1}{p_Y(y=0)}+p_Y(y=1)\log\frac{1}{p_Y(y=1)}\\&=\frac{1}{4}\log 4+\frac{3}{4}\log\frac{4}{3}=\log 2=2-\frac{3}{4}\log 3\end{aligned}$$

(b)

$$\begin{aligned}H(X|Y=0)&=p_{X|Y}(x=0|y=0)\log\frac{1}{p_{X|Y}(x=0|y=0)}+p_{X|Y}(x=1|Y=0)\log\frac{1}{p_{X|Y}(x=1|y=0)}\\&=0\end{aligned}$$

$$\begin{aligned}H(X|Y=1)&=p_{X|Y}(x=0|y=1)\log\frac{1}{p_{X|Y}(x=0|y=1)}+p_{X|Y}(x=1|Y=1)\log\frac{1}{p_{X|Y}(x=1|y=1)}\\&=\frac{1}{3}\log 3+\frac{2}{3}\log\frac{3}{2}=\log 3-\frac{2}{3}\end{aligned}$$

$$\begin{aligned}H(Y|X=0)&=p_{Y|X}(y=0|x=0)\log\frac{1}{p_{Y|X}(y=0|x=0)}+p_{Y|X}(y=1|x=0)\log\frac{1}{p_Y(y=1|x=0)}\\&=\frac{1}{2}\log 2+\frac{1}{2}\log 2=\log 2=1\end{aligned}$$

$$\begin{aligned}H(Y|X=1)&=p_{Y|X}(y=0|x=1)\log\frac{1}{p_{Y|X}(y=0|x=1)}+p_{Y|X}(y=1|x=1)\log\frac{1}{p_Y(y=1|x=1)}\\&=0\end{aligned}$$

(c)

$$\begin{aligned} H(X|Y) &= p_Y(y=0)H(X|Y=0) + p_Y(y=1)H(X|Y=1) \\ &= \frac{1}{4}0 + \frac{3}{4}\left(\log 3 - \frac{2}{3}\right) = \frac{3}{4}\log 3 - \frac{1}{2} \end{aligned}$$

$$\begin{aligned} H(Y|X) &= p_X(x=0)H(Y|X=0) + p_X(x=1)H(Y|X=1) \\ &= \frac{1}{2}1 + \frac{1}{2}0 = \frac{1}{2} \end{aligned}$$

(d)

$$\begin{aligned} H(X,Y) = {} & p_{XY}(x=0,y=0)\frac{1}{p_{XY}(x=0,y=0)} + p_{XY}(x=0,y=1)\frac{1}{p_{XY}(x=0,y=1)} \\ & + p_{XY}(x=1,y=0)\frac{1}{p_{XY}(x=1,y=0)} + p_{XY}(x=1,y=1)\frac{1}{p_{XY}(x=1,y=1)} \\ & = \frac{1}{4}\log 4 + \frac{1}{4}\log 4 + 0\log\frac{1}{0} + \frac{1}{2}\log 2 \\ & = \frac{3}{2} \end{aligned}$$

(e)

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) = 1 - \left(\frac{3}{4}\log 3 - \frac{1}{2}\right) = \frac{3}{2} - \frac{3}{4}\log 3 \\ &= H(Y) - H(Y|X) = \left(2 - \frac{3}{4}\log 3\right) - \frac{1}{2} = \frac{3}{2} - \frac{3}{4}\log 3 \end{aligned}$$

(e) Part **(b)** asks us to compute entropy of Random variable X conditioned on the *event* $Y = y$. Conditioning on a random variable will always reduce entropy. In particular, it means that the average of $H(X|Y = y)$ can at most be equal to $H(X)$. Thus, there is no violation (but perhaps a mild misdirection/trick question?).

Problem 2: Data Processing Inequality

(a) Show that $H(X|Y) = H(X|Y,Z)$.

$$\begin{aligned} H(X|Y,Z) &= \sum_{x,y,z} p_{XYZ}(x,y,z)\log\frac{1}{p_{X|YZ}(x|y,z)} \\ &= \sum_{x,y,z} p_{XYZ}(x,y,z)\log\frac{1}{p_{X|YZ}(x|y)} \\ &= \sum_{x,y}\sum_{z} p_{XYZ}(x,y,z)\log\frac{1}{p_{X|Y}(x|y)} \\ &= \sum_{x,y}\log\frac{1}{p_{X|Y}(x|y)}\sum_{z} p_{XYZ}(x,y,z) \\ &= \sum_{x,y} p_{XY}(x,y)\log\frac{1}{p_{X|Y}(x|y)} \\ &= H(X|Y) \end{aligned}$$

Or, if you would like to start with the other end,

$$\begin{aligned}H(X|Y) &= \sum_{x,y} p_{XY}(x,y)\log\frac{1}{p_{X|Y}(x|y)}\\ &= \sum_{x,y,z} p_{XYZ}(x,y,z)\log\frac{1}{p_{X|Y}(x|y)}\\ &= \sum_{x,y,z} p_{XYZ}(x,y,z)\log\frac{1}{p_{X|YZ}(x|y,z)}\\ &= H(X|Y,Z).\end{aligned}$$

(b) Show that $H(X|Y) \leq H(X|Z)$.

$$\begin{aligned}H(X|Y) &= H(X|Y,Z)\\ &\leq H(X|Y,Z) \text{ (since conditioning reduces entropy)}.\end{aligned}$$

(c) Show that $I(X;Y) \geq I(X;Z)$.

$$\begin{aligned}I(X;Y) &= H(X) - H(X|Y)\\ &\geq H(X) - H(X|Z) \text{ (from part (b))}\\ &= I(X;Z).\end{aligned}$$

(d) Show that $I(X;Z|Y) = 0$.

$$\begin{aligned}I(X;Z|Y) &= \mathbb{E}\left[\log\frac{p_{XZ|Y}(X,Z|Y)}{p_{X|Y}(X|Y)p_{Z|Y}(Z|Y)}\right]\\ &= \mathbb{E}\left[\log\frac{p_{X|Y}(X|Y)p_{Z|Y,X}(Z|X,Y)}{p_{X|Y}(X|Y)p_{Z|Y}(Z|Y)}\right]\\ &= \mathbb{E}\left[\log\frac{p_{X|Y}(X|Y)p_{Z|Y}(Z|Y)}{p_{X|Y}(X|Y)p_{Z|Y}(Z|Y)}\right]\\ &= \mathbb{E}[\log 1] = 0.\end{aligned}$$

Problem 3: Two Looks

(a) Given $I(X;Y_1) = I(X;Y_1) = 0$, does it follow that $I(X;Y_1,Y_2) = 0$?

No. Consider $Y_1 \sim \mathcal{B}(\frac{1}{2}), Y_2 \sim \mathcal{B}(\frac{1}{2})$, with alphabet $\{0,1\}$. Let $X = Y_1 + Y_2 \mod 2$. Verify $I(X;Y_1) = I(X;Y_2) = 0$, and $I(X;Y_1,Y_2) \neq 0$.

(b) Given $I(X;Y_1) = I(X;Y_1) = 0$, does it follow that $I(Y_1;Y_2) = 0$?

No. Consider $Y_1 \sim \mathcal{B}(\frac{1}{2})$, with alphabet $\{0,1\}$, $Y_2 = Y_1$. Let $X = Y_1 + Y_2 \mod 2$. Verify $I(X;Y_1) = I(X;Y_2) = 0$, and $I(Y_1;Y_2) \neq 0$.

Problem 4

(a) Using any one of the many techniques learnt in class or as given in the text book, one can derive the chain rule as given by

$$I(X^n;Y^n) = \sum_i I(X^n;Y_i|Y^{i-1}),$$

and therefore the proof is omitted herein.

(b) Given $I(X^n \to Y^n) = \sum_i I(X^i;Y_i|Y^{i-1})$. Show that $I(X^n;Y^n) \geq I(X^n \to Y^n)$.

$$\begin{aligned}I(X^n;Y^n) &= \sum_i I(X^n;Y_i|Y^{i-1})\\ &= \sum_i I(X^i;Y_i|Y^{i-1}) + I(X^n_{i+1};Y_i|X^i,Y^{i-1}) \text{ (by chain rule)}\\ &\leq \sum_i I(X^i;Y_i|Y^{i-1}) \text{ (by non negativity of mutual information)}\\ &= I(X^n \to Y^n).\end{aligned}$$

(c) Preservation Law There are many ways to show this. One such method is illustrated below. From (b), it is sufficient to prove that $\sum_i I(X_{i+1}^n; Y_i | X^i, Y^{i-1}) = I(Y^{n-1} \to X^n)$

$$
\begin{aligned}
\sum_{i=1}^{n} I(X_{i+1}^n; Y_i | X^i, Y^{i-1}) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} I(X_j; Y_i | X^{j-1}, Y^{i-1}) \\
&= \sum_{j=1}^{n} \sum_{i=1}^{j-1} I(X_j; Y_i | X^{j-1}, Y^{i-1}) \text{ (changing the order of sums)} \\
&= \sum_{j=1}^{n} I(Y^{i-1}; X_j | X^{j-1}) \\
&= I(Y^{n-1} \to X^n).
\end{aligned}
$$

Question 5

9) (*AEP*).

a) Since the $X_1, X_2, \ldots, X_n$ are i.i.d., so are $q(X_1), q(X_2), \ldots, q(X_n)$, and hence we can apply the strong law of large numbers to obtain

$$\begin{aligned}
\lim -\frac{1}{n}\log q(X_1, X_2, \ldots, X_n) &= \lim -\frac{1}{n}\sum \log q(X_i) && (158)\\
&= -E(\log q(X)) \text{ w.p. } 1 && (159)\\
&= -\sum p(x)\log q(x) && (160)\\
&= \sum p(x)\log\frac{p(x)}{q(x)} - \sum p(x)\log p(x) && (161)\\
&= D(\mathbf{p}||\mathbf{q}) + H(\mathbf{p}). && (162)
\end{aligned}$$

b) Again, by the strong law of large numbers,

$$\begin{aligned}
\lim -\frac{1}{n}\log\frac{q(X_1, X_2, \ldots, X_n)}{p(X_1, X_2, \ldots, X_n)} &= \lim -\frac{1}{n}\sum \log\frac{q(X_i)}{p(X_i)} && (163)\\
&= -E(\log\frac{q(X)}{p(X)}) \text{ w.p. } 1 && (164)\\
&= -\sum p(x)\log\frac{q(x)}{p(x)} && (165)\\
&= \sum p(x)\log\frac{p(x)}{q(x)} && (166)\\
&= D(\mathbf{p}||\mathbf{q}). && (167)
\end{aligned}$$

11.

a) Let A^c denote the complement of A. Then

$$P(A^c \cup B^c) \le P(A^c) + P(B^c). \qquad (168)$$

Since $P(A) \ge 1 - \epsilon_1$, $P(A^c) \le \epsilon_1$. Similarly, $P(B^c) \le \epsilon_2$. Hence

$$\begin{aligned}
P(A \cap B) &= 1 - P(A^c \cup B^c) && (169)\\
&\ge 1 - P(A^c) - P(B^c) && (170)\\
&\ge 1 - \epsilon_1 - \epsilon_2. && (171)
\end{aligned}$$

b) To complete the proof, we have the following chain of inequalities

$$\begin{aligned}
1 - \epsilon - \delta &\stackrel{(a)}{\le} \Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) && (172)\\
&\stackrel{(b)}{=} \sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} p(x^n) && (173)\\
&\stackrel{(c)}{\le} \sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} 2^{-n(H-\epsilon)} && (174)\\
&\stackrel{(d)}{=} |A_\epsilon^{(n)} \cap B_\delta^{(n)}| 2^{-n(H-\epsilon)} && (175)\\
&\stackrel{(e)}{\le} |B_\delta^{(n)}| 2^{-n(H-\epsilon)}. && (176)
\end{aligned}$$

where (a) follows from the previous part, (b) follows by definition of probability of a set, (c) follows from the fact that the probability of elements of the typical set are bounded by $2^{-n(H-\epsilon)}$, (d) from the definition of $|A_\epsilon^{(n)} \cap B_\delta^{(n)}|$ as the cardinality of the set $A_\epsilon^{(n)} \cap B_\delta^{(n)}$, and (e) from the fact that $A_\epsilon^{(n)} \cap B_\delta^{(n)} \subseteq B_\delta^{(n)}$.

13) a) $H(X) = -0.6 \log 0.6 - 0.4 \log 0.4 = 0.97095$ bits.

b) By definition, $A_\epsilon^{(n)}$ for $\epsilon = 0.1$ is the set of sequences such that $-\frac{1}{n} \log p(x^n)$ lies in the range $(H(X) - \epsilon, H(X) + \epsilon)$, i.e., in the range (0.87095, 1.07095). Examining the last column of the table, it is easy to see that the typical set is the set of all sequences with the number of ones lying between 11 and 19. The probability of the typical set can be calculated from cumulative probability column. The probability that the number of 1's lies between 11 and 19 is equal to $F(19) - F(10) = 0.970638 - 0.034392 = 0.936246$. Note that this is greater than $1 - \epsilon$, i.e., the n is large enough for the probability of the typical set to be greater than $1 - \epsilon$.

The number of elements in the typical set can be found using the third column.

$$|A_\epsilon^{(n)}| = \sum_{k=11}^{19} \binom{n}{k} = \sum_{k=0}^{19} \binom{n}{k} - \sum_{k=0}^{10} \binom{n}{k} = 33486026 - 7119516 = 26366510. \tag{177}$$

Note that the upper and lower bounds for the size of the $A_\epsilon^{(n)}$ can be calculated as $2^{n(H+\epsilon)} = 2^{25(0.97095+0.1)} = 2^{26.77} = 1.147365 \times 10^8$, and $(1-\epsilon)2^{n(H-\epsilon)} = 0.9 \times 2^{25(0.97095-0.1)} = 0.9 \times 2^{21.9875} = 3742308$. Both bounds are very loose!

c) To find the smallest set $B_\delta^{(n)}$ of probability 0.9, we can imagine that we are filling a bag with pieces such that we want to reach a certain weight with the minimum number of pieces. To minimize the number of pieces that we use, we should use the largest possible pieces. In this case, it corresponds to using the sequences with the highest probability.

Thus we keep putting the high probability sequences into this set until we reach a total probability of 0.9. Looking at the fourth column of the table, it is clear that the probability of a sequence increases monotonically with k. Thus the set consists of sequences of $k = 25, 24, \ldots,$ until we have a total probability 0.9.

Using the cumulative probability column, it follows that the set $B_\delta^{(n)}$ consist of sequences with $k \geq 13$ and some sequences with $k = 12$. The sequences with $k \geq 13$ provide a total probability of $1 - 0.153768 = 0.846232$ to the set $B_\delta^{(n)}$. The remaining probability of $0.9 - 0.846232 = 0.053768$ should come from sequences with $k = 12$. The number of such sequences needed to fill this probability is at least $0.053768/p(x^n) = 0.053768/1.460813 \times 10^{-8} = 3680690.1$, which we round up to 3680691. Thus the smallest set with probability 0.9 has $33554432 - 16777216 + 3680691 = 20457907$ sequences. Note that the set $B_\delta^{(n)}$ is not uniquely defined - it could include any 3680691 sequences with $k = 12$. However, the size of the smallest set is well defined.

d) The intersection of the sets $A_\epsilon^{(n)}$ and $B_\delta^{(n)}$ in parts (b) and (c) consists of all sequences with k between 13 and 19, and 3680691 sequences with $k = 12$. The probability of this intersection $= 0.970638 - 0.153768 + 0.053768 = 0.870638$, and the size of this intersection $= 33486026 - 16777216 + 3680691 = 20389501$.