Assignment 1

Ge Yao, alexgecontrol@qq.com

Problem 1

(1) Since X has the following marginal probability distribution:

X	1	2	3	4	5
$p(x)$	0.200	0.200	0.200	0.200	0.200

$$\therefore H_2(X) = \sum_x -log_2(x)p(x) = 2.3219$$

(2) Since Y has the following marginal probability distribution:

Y	1	2	3	4	5
$p(y)$	0.200	0.200	0.200	0.200	0.200

$$\therefore H_2(Y) = \sum_y -log_2(y)p(y) = 2.3219$$

(3) Since X has the following conditional entropy:

Y	1	2	3	4	5
$H_2(X\|Y=y)$	1.5219	1.3710	1.9219	1.9219	1.3710

$$\therefore H_2(X|Y) = \sum_y H_2(X|Y=y)p(y) = 1.6215$$

(4) Since Y has the following conditional entropy:

X	1	2	3	4	5
$H_2(Y\|X=x)$	2.3219	1.5219	1.9219	0.9710	1.3710

$$\therefore H_2(Y|X) = \sum_x H_2(Y|X=x)p(x) = 1.6215$$

(5) Since the joint entropy of X and Y is:

$$H_2(X,Y) = \sum_{x,y} -log_2(x,y)p(x,y) = 3.9435$$

$$\therefore I_2(X;Y) = H_2(X) + H_2(Y) - H_2(X,Y) = 0.7004$$

Problem 2

(1) Proposition 2.8

Necessity

$$\because p(x_1, x_2, \dots, x_n) = p(x_1, x_2)p(x_3|x_2) \dots p(x_n|x_{n-1})$$

$$\therefore \sum_{x_n} p(x_1, x_2, \dots, x_n) = p(x_1, x_2) \dots p(x_{n-1}|x_{n-2}) \sum_{x_n} p(x_n|x_{n-1})$$

$$\therefore p(x_1, x_2, \dots, x_{n-1}) = p(x_1, x_2) \dots p(x_{n-1}|x_{n-2})$$

Following the same method we can also have:

$$p(x_1, x_2, \dots, x_{n-2}) = p(x_1, x_2) \dots p(x_{n-2}|x_{n-1})$$

$$\dots$$

$$p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3|x_2)$$

The last equation is equivalent to:

$$p(x_1, x_3|x_2) = p(x_1|x_2)p(x_3|x_2)$$

So we have:

$$X_1 \perp X_3|X_2 \leftrightarrow X_1 \to X_2 \to X_3$$

Substitute

$$p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3|x_2)$$

into

$$p(x_1, x_2, x_3, x_4) = p(x_1, x_2)p(x_3|x_2)p(x_4|x_3)$$

We also have:

$$p(x_1, x_2, x_3, x_4) = p(x_1, x_2, x_3)p(x_4|x_3)$$

which is equivalent to:

$$p(x_1, x_2, x_4|x_3) = p(x_1, x_2|x_3)p(x_4|x_3)$$

So we also have:

$$(X_1, X_2) \perp X_4|X_3 \leftrightarrow (X_1, X_2) \to X_3 \to X_4$$

Following the same method we also have:

$$(X_1, X_2, X_3) \perp X_5|X_4 \leftrightarrow (X_1, X_2, X_3) \to X_4 \to X_5$$

$$\dots$$

$$(X_1, X_2, \dots, X_{n-2}) \perp X_n|X_{n-1} \leftrightarrow (X_1, X_2, \dots, X_{n-2}) \to X_{n-1} \to X_n$$

Thus the necessity has been proved.

Sufficiency

$$\because X_1 \to X_2 \to X_3 \leftrightarrow p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3|x_2)$$

$$(X_1, X_2) \to X_3 \to X_4 \leftrightarrow p(x_1, x_2, x_3, x_4) = p(x_1, x_2, x_3)p(x_4|x_3)$$

Substitute the former into the latter we have:

$$p(x_1, x_2, x_3, x_4) = p(x_1, x_2)p(x_3|x_2)p(x_4|x_3)$$

which indicates that:

$$X_1 \to X_2 \to X_3 \to X_4$$

Following the same method we also have:

$$X_1 \to X_2 \to X_3 \to X_4 \to X_5$$

$$\dots$$

$$X_1 \to X_2 \to \dots \to X_n$$

Thus the sufficiency has been proved. $Q.E.D.$

(2) Proposition 2.9

Necessity

The necessity part of the proof follows immediately from the definition of Markov chain.

Sufficiency

$$\because p(x_1, x_2, \dots, x_n) = f_1(x_1, x_2) \dots f_{n-1}(x_{n-1}, x_n)$$

$$\therefore p(x_1, x_2, \dots, x_{n-1}) = f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1}) \left(\sum_{x_n} f_{n-1}(x_{n-1}, x_n) \right)$$

$$\therefore p(x_{n-1}, x_n) = \left(\sum_{x_1, \dots, x_{n-2}} f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1}) \right) f_{n-1}(x_{n-1}, x_n)$$

$$\therefore p(x_{n-1}) = \left(\sum_{x_1, \dots, x_{n-2}} f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1}) \right) \left(\sum_{x_n} f_{n-1}(x_{n-1}, x_n) \right)$$

$$\therefore p(x_1, \dots, x_{n-2}, x_n | x_{n-1}) = \frac{f_1(x_1, x_2) \dots f_{n-1}(x_{n-1}, x_n)}{\left(\sum_{x_1, \dots, x_{n-2}} f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1})\right)\left(\sum_{x_n} f_{n-1}(x_{n-1}, x_n)\right)}$$

$$= \frac{f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1})\left(\sum_{x_n} f_{n-1}(x_{n-1}, x_n)\right)\left(\sum_{x_1, \dots, x_{n-2}} f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1})\right) f_{n-1}(x_{n-1}, x_n)}{\left[\left(\sum_{x_1, \dots, x_{n-2}} f_1(x_1, x_2) \dots f_{n-2}(x_{n-2}, x_{n-1})\right)\left(\sum_{x_n} f_{n-1}(x_{n-1}, x_n)\right)\right]^2}$$

$$= \frac{p(x_1, x_2, \dots, x_{n-1}) p(x_{n-1}, x_n)}{p^2(x_{n-1})} = p(x_1, \dots, x_{n-2} | x_{n-1}) p(x_n | x_{n-1})$$

$$\therefore (X_1, X_2, \dots, X_{n-2}) \perp X_n | X_{n-1} \leftrightarrow (X_1, X_2, \dots, X_{n-2}) \to X_{n-1} \to X_n$$

Following the same method we also have:

$$(X_1, X_2, \dots, X_{n-3}) \perp X_{n-1} | X_{n-2} \leftrightarrow (X_1, X_2, \dots, X_{n-3}) \to X_{n-2} \to X_{n-1}$$

$$\dots$$

$$X_1 \perp X_3 | X_2 \leftrightarrow X_1 \to X_2 \to X_3$$

Thus the sufficiency has been proved. $Q.E.D.$

(3) Proposition 2.19

$$\because I(X;Y) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_{x,y} p(x,y) \cdot -logp(x) - \sum_{x,y} p(x,y) \cdot -logp(x|y) = \boldsymbol{H(X) - H(X|Y)}$$

$$= \sum_{x,y} p(x,y) \cdot -logp(y) - \sum_{x,y} p(x,y) \cdot -logp(y|x) = \boldsymbol{H(Y) - H(Y|X)}$$

$$= \sum_{x,y} p(x,y) \cdot -logp(x) + \sum_{x,y} p(x,y) \cdot -logp(y) - \sum_{x,y} p(x,y) \cdot -logp(x,y) = \boldsymbol{H(X) + H(Y) - H(X,Y)}$$

$$\therefore Q.E.D.$$

(4) Proposition 2.21

$$\because I(X;X|Z) = \sum_{z} p(z) \sum_{x} p(x|z) log \frac{p(x|z)}{p^2(x|z)}$$

$$= \sum_{z} p(z) \sum_{x} p(x|z) \cdot -logp(x|z) = \sum_{z} p(z) \sum_{x} H(X|Z=z) = \boldsymbol{H(X|Z)}$$

$$\therefore Q.E.D.$$

(5) Proposition 2.22

$$\because I(X;Y|Z) = \sum_{z} p(z) \sum_{x,y} p(x,y|z) log \frac{p(x,y|z)}{p(x|z)p(y|z)}$$

$$= \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(x|z) - \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(x|y,z)$$

$$= \sum_{x,y,z} p(x,y,z) \cdot -logp(x|z) - \sum_{x,y,z} p(x,y,z) \cdot -logp(x|y,z) = \boldsymbol{H(X|Z) - H(X|Y,Z)}$$

$$= \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(y|z) - \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(y|x,z)$$

$$= \sum_{x,y,z} p(x,y,z) \cdot -logp(y|z) - \sum_{x,y,z} p(x,y,z) \cdot -logp(y|x,z) = \boldsymbol{H(Y|Z) - H(Y|X,Z)}$$

$$= \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(x|z) + \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(y|z) - \sum_{z} p(z) \sum_{x,y} p(x,y|z) \cdot -logp(x,y|z)$$

$$= \sum_{x,y,z} p(x,y,z) \cdot -logp(x|z) + \sum_{x,y,z} p(x,y,z) \cdot -logp(y|z) - \sum_{x,y,z} p(x,y,z) \cdot -logp(x,y|z) = \boldsymbol{H(X|Z) + H(Y|Z) - H(X,Y|Z)}$$

$$\therefore Q.E.D.$$

Problem 3

Here is a case from ***Probability and Statistics*** (China University of Science and Technology Press 2009) by Prof. Xiru Chen on page 87:

Put four identical balls with tags ‘1’, ‘2’, ‘3’ and ‘1, 2 and 3’ respectively inside a bag then draw a ball from the bag. Define the following accidents:

$$A_i = \{Number\ 'i'\ is\ on\ the\ tag\ of\ the\ ball\}, i = 1,2,3$$

Then we have:

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$$

$$P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = \frac{1}{4}$$

Since:

$$P(A_iA_j) = P(A_i)P(A_j), i \neq j, 1 \leq i,j \leq 3$$

So **the accidents are pairwise independent**. However:

$$P(A_1A_2A_3) = \frac{1}{4}$$

$$P(A_1)P(A_2)P(A_3) = \frac{1}{8}$$

So **the accidents are NOT mutually independent**. To sum up, pairwise independency does not equal to mutual independence.

Problem 4

First, it’s obvious that Definition 2.4 satisfies **non-negativity** and **sigma-additivity**. Since we also have:

$$\sum_{\mathrm{x,y,z}} \mathrm{p}(x,y,z) = \sum_{x,y,z} \mathrm{p}(x,y)p(z|y) = \sum_{y,z} p(z|y) \sum_{x} p(x,y)$$

$$= \sum_{y,z} p(z|y)p(y) = \sum_{z} \sum_{y} p(z|y)p(y) = \sum_{z} p(z) = 1$$

So Definition 2.4 also satisfies unitarity so Definition 2.4 is a probability distribution.

Problem 5

According to the definition of Lebesgue integral, integration on the support set of the integrand is equivalent to integration on the natural domain which is the whole alphabet for probability.
Since probability is in essence an abstract measure and the expectation is in essence an abstract integration in respect to the given measure, so expectation taken over the supporting set is equivalent to that taken over whole alphabet.

Problem 8

In the case of finite alphabet, we can use the following corollary from matrix analysis to complete the proof.

Lemma: Let $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ ($\|x\|_t \triangleq \left(\sum_{i=1}^{N} x_i^t\right)^{1/t}, t > 0$) be given norms on a finite-dimensional real vector space V. Then there exist finite positive constants C_M and C_N such that $C_M\|x\|_\alpha \leq \|x\|_\beta \leq C_N\|x\|_\alpha$ for all $x \in$ V.

The detailed proof of this lemma can be found in ***Matrix Analysis,*** Second Edition by Roger A. Horn and Charles R. Johnson.
The implication of this lemma is that **convergences in respect to any two norms are equivalent**. Using the above lemma on variational distance and L2 distance, we have:

$$\exists \mathrm{C}_M^1, \mathrm{C}_N^1, \mathrm{C}_M^2, \mathrm{C}_M^2$$
$$\mathrm{C}_M^1\|p_k - p\|_1 \leq \|p_k - p\|_2 \leq \mathrm{C}_N^1\|p_k - p\|_1, \forall k \in N_+$$
$$\mathrm{C}_M^2\|p_k - p\|_2 \leq \|p_k - p\|_1 \leq \mathrm{C}_N^2\|p_k - p\|_2, \forall k \in N_+$$

So we finally have:

$$\lim_{k\to\infty} \|p_k - p\|_2 = 0 \leftrightarrow \lim_{k\to\infty} \|p_k - p\|_1$$

That is, the variational distance and L2 distance are equivalent.