CSE 250A HW1

Jiping Lin A15058075

1. (a) $LHS = P(X,Y|E) = \frac{P(X,Y,E)}{P(E)}$; $RHS = P(X|Y,E)P(Y|E)$

$= \frac{P(X,Y,E)}{P(Y,E)} P(Y|E)$

$= \frac{P(X,Y,E)}{P(Y|E)P(E)} P(Y|E)$

$= \frac{P(X,Y,E)}{P(E)} = LHS$

(b) $P(X|Y,E)P(Y|E)$

$= P(X,Y|E)$ (by (a))

$= P(Y,X|E) = P(Y|X,E)P(X,E)$

So $P(X|Y,E) = \frac{P(Y|X,E)P(X,E)}{P(Y|E)}$

(c) $LHS = P(X|E) = \frac{P(X,E)}{P(E)}$, $RHS = \sum_y P(X,Y=y|E) = \sum_y \frac{P(X,Y=y,E)}{P(E)} = \frac{\sum_y P(X,Y=y,E)}{P(E)}$

So $LHS = RHS \Leftrightarrow P(X,E) = \sum_y P(X,Y=y,E)$,

where $\sum_y P(X,Y=y,E) = \sum_y P(X)P(E|X)P(Y=y|X,E)$ ⌜Product Rule⌟

$= P(X)P(E|X)\sum_y P(Y=y|X,E)$

$= P(X)P(E|X)$ ⌜Since $\sum_y P(Y=y|X,E)$ sums over all y given X,E so it is 1⌟

$= P(X,E)$.

2. (1)⇒(2):

(1) ⇒ $P(X|E) = \frac{P(X,Y|E)}{P(Y|E)} = \frac{P(X|Y,E)P(Y|E)}{P(Y|E)}$ ⌜by 1(a)⌟

$= P(X|Y,E)$

(2)⇒(3):

(2) ⇒ $P(X|Y,E)P(Y|E) = P(X|E)P(Y|E)$

$= P(X,Y|E)$ ⌜by 1(a)⌟

$= P(Y,X|E) = P(Y|X,E)P(X|E)$

⇒ $P(X|E)P(Y|E) = P(Y|X,E)P(X|E) = P(Y|E)P(Y|X,E)$

(3)⇒(1):

(3) ⇒ $P(Y|X,E)P(X|E) = P(Y|E)P(X|E)$

$= P(Y,X|E)$ ⌜by 1(a)⌟ $= P(X,Y|E)$

3. (a) Let X denote whether a driver got pulled over,

Y denote whether the driver is drunk,

Z denote whether the driver is speeding.

then $P(X=1) < P(X=1|\underbrace{Y=1}_{\text{Drunk}}) < P(X=1|\underbrace{Y=1, Z=1}_{\text{Drunk and speed}})$

(b) Keep X, Y same, change Z to whether he is driving in local

Assuming more police on freeway, we have

$P(X=1|Y=1,Z=1) < P(X=1|Y=1)$

(c) Suppose Abby and Bob live in the same town but do not know each other and have no interaction but working in the same place, let X be Abby arrives late for work, Y be Bob arrives late for work.

Then X, Y are not independent since if $X=1$ (Abby late), the cause for it might also affect Y (e.g. traffic or weather), let Z be whether the work is remote, then if $Z=1$, then X and Y are conditionally independent given $Z=1$, since $Z=1$ rules out all possibilities that can cause both $X=1$ and $Y=1$.

4. (a)

$$D \quad P(D=0)=0.99$$
$$\downarrow \quad P(D=1)=0.01$$
$$T$$

D	P(T=1|D)
0	0.05
1	0.9

(b) We want $P(D=0 \mid T=0)$

Note $P(T=0)=P(T=0,D=0)+P(T=0,D=1)$

$$=P(T=0\mid D=0)P(D=0)+P(T=0\mid D=1)P(D=1)$$
$$=(1-0.05)0.99+(1-0.9)0.01$$
$$=0.9415,$$

By Baye's Rule, we have

$$P(D=0\mid T=0)=\frac{P(T=0\mid D=0)P(D=0)}{P(T=0)}$$
$$=\frac{(1-0.05)0.99}{0.9415}$$
$$=99.89\%$$

(c) We want $P(D=1\mid T=1)$,

$$P(D=1\mid T=1)=\frac{P(T=1\mid D=1)P(D=1)}{P(T=1)}$$
$$=\frac{0.9\times 0.01}{1-0.9415}=15.38\%$$

5. (a) Write $L(p_1,\dots,p_n,\lambda)=-\sum_{i=1}^{n}p_i\log p_i+\lambda\left(\sum_{i=1}^{n}p_i-1\right)$,

then we want

$$\frac{\partial L}{\partial p_i}=-\left(p_i\cdot\frac{1}{p_i}+\log p_i\right)+\lambda=0,\quad \forall i=1,\dots,n,$$

and $\frac{\partial L}{\partial \lambda}=\sum_{i=1}^{n}p_i-1=0 \iff \sum_{i=1}^{n}p_i=1$

we have $\lambda-1-\log p_i=0 \Rightarrow \log p_i=\lambda-1 \ \forall i=1,\dots,n$.

this implies $p_1=\dots=p_n$, since $\sum_{i=1}^{n}p_i=1$,

then we must have $p_i=\frac{1}{n}$, $\forall i$.

(b) We first prove for any X_1, X_2, $X_1 \perp\!\!\!\perp X_2$, $H(X_1 X_2) = H(X_1) + H(X_2)$:

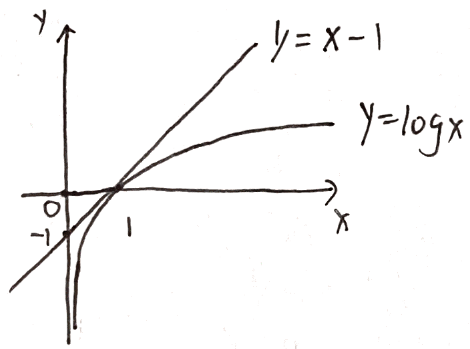
We have
$$
\begin{aligned}
H(X_1, X_2) &= -\sum_{x_1}\sum_{x_2} P(X_1, X_2)\log P(X_1, X_2) \\
&= -\sum_{x_1}\sum_{x_2} P(X_1)P(X_2)(\log P(X_1) + \log P(X_2)) \\
&= -\sum_{x_1}\sum_{x_2} P(X_1)P(X_2)\log P(X_1) - \sum_{x_1}\sum_{x_2} P(X_1)P(X_2)\log P(X_2) \\
&= -\sum_{x_1} P(X_1)\log P(X_1)\sum_{x_2} P(X_2) - \sum_{x_1} P(X_1)\sum_{x_2} P(X_2)\log P(X_2) \\
&= -\sum_{x_1} P(X_1)\log P(X_1) - \sum_{x_2} P(X_2)\log P(X_2) \\
&= H(X_1) + H(X_2).
\end{aligned}
$$

If we have $X_1, \cdots, X_n$ which are mutually independent,

then $H(X_1 \cdots X_n) = H((X_1 \cdots X_{n-1})X_n) = H(X_1 \cdots X_{n-1}) + H(X_n)$

$= \cdots = H(X_1) + \cdots + H(X_n)$ by induction.

6.(a)



From the graph we can see that

$x - 1 \geq \log x$, $\forall x > 0$, and

$x - 1 = \log x \iff x = 1$.

Alternatively,

$\frac{\partial}{\partial x}[\log x - (x-1)] = \frac{1}{x} - 1 = 0 \iff x = 1$,

and $\frac{\partial^2}{\partial x}\Big|_{x=1} = -\frac{1}{x^2}\Big|_{x=1} = -1 < 0$,

so $x = 1$ is the global min of $\log x - (x-1)$,

and since $x = 1 \iff \log x - (x-1) = 0$,

then $\log x \leq x - 1$ $\forall x$ and "=" holds only when $x = 1$.

(b)
$$
\begin{aligned}
KL(p, q) &= \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \\
&= \sum_i p_i \log\left(\left(\frac{q_i}{p_i}\right)^{-1}\right) \\
&= -\sum_i p_i \log\left(\frac{q_i}{p_i}\right) \quad \left\ulcorner \log\frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1 \quad (*) \right\lrcorner \\
&\geq -\sum_i p_i\left(\frac{q_i}{p_i} - 1\right) \quad (**) \\
&= -\sum_i (q_i - p_i) = -\sum_i q_i + \sum_i p_i = -1 + 1 = 0
\end{aligned}
$$

For equality holds in (**), we must have equality holds in (*).

By (a), this is true iff $\frac{q_i}{p_i} = 1$, i.e. $p_i = q_i$, $\forall i$.

(c) $KL(p, q) = -\sum_i p_i \log\left(\frac{q_i}{p_i}\right) = -\sum_i p_i 2\log\sqrt{\frac{q_i}{p_i}} = -2\sum_i p_i \log\sqrt{\frac{q_i}{p_i}} \geq -2\sum_i p_i\left(\sqrt{\frac{q_i}{p_i}} - 1\right)$

then $KL(p, q) \geq -2\sum_i \sqrt{p_i q_i} - p_i = -2\sum_i \sqrt{p_i q_i} + 2\sum_i p_i = 2 - 2\sum_i \sqrt{p_i q_i}$,

$\sum_i (\sqrt{p_i} - \sqrt{q_i})^2 = \sum_i (p_i + q_i - 2\sqrt{p_i q_i}) = \sum_i p_i + \sum_i q_i - 2\sum_i \sqrt{p_i q_i} = 2 - 2\sum_i \sqrt{p_i q_i}$,

So $KL(p, q) \geq \sum_i (\sqrt{p_i} - \sqrt{q_i})^2$

(d) Let $p_1 = 0.8 \quad q_1 = 0.4$,
$p_2 = 0.2 \quad q_2 = 0.6$

then $KL(P, q) = 0.8 \log 2 + 0.2 \log \frac{1}{3} = 0.335$

$KL(q, P) = 0.4 \log \frac{1}{2} + 0.6 \log 3 = 0.382$

7. (a) Fix x to be a possible value of X,

then write $f_x(Y) = \sum_y P(x, y) \log\left(\frac{P(X,Y)}{P(X)P(Y)}\right)$,

then $f_x(Y) = \sum_y P(y \mid X=x) \log\left(\frac{P(Y \mid X=x)}{c P(Y)}\right)$ $\ulcorner P(x) = c > 0 \lrcorner$,

$\geqslant \sum_y P(y \mid X=x) \log\left(\frac{P(Y \mid X=x)}{P(Y)}\right)$ since $c \leq 1$,

$= KL(Y \mid X=x, Y) \geqslant 0$ by 6(b)

then $I(X, Y) = \sum_x f_x(Y) \geqslant 0$.

(b) If $I(X, Y) = 0$, then $f_x(Y) = 0$, $\forall x$, that means $KL(Y \mid X=x, Y) = 0$, $\forall x$. By 6(b), $KL(Y \mid X=x, Y) = 0 \Leftrightarrow (Y \mid X=x) = Y$, $\forall x$.

So this implies $X \perp\!\!\!\perp Y$.

If $X \perp\!\!\!\perp Y$, then $P(X, Y) = P(X)P(Y) \Rightarrow I(X, Y) = \sum_x \sum_y P(x, y) \log 1 = 0$.

8. (a) First BN implies conditional independence of Y, Z given X; while in second BN, Y, Z are dependent given (or not given) X.

(b) Second BN implies marginal independence of X, Z, but in third BN, X, Z are also marginal independent.
So Second BN doesn't contain the desired independence.

(c) Third BN implies marginal independence of X, Z, while in First BN, X, Z are dependent.

9. (a) Most: three seven eight would about,
their which after first fifty,
other forty years there sixty.

Least: troup ottis mapco caixa bosak,
yalom tocor serna paxon niaid,
foamy fabri cleft ccair

these make sense since top part words are frequently seen, and I know nothing of the bot part words

(b) Missing parts are

E, 0.5394
O, 0.5340
E, 0.7715
E, 0.7127
R, 0.7454.