1.1

(a) Prove $P(X,Y|E) = P(X|Y,E)P(Y|E)$ ...

Product Rule : $P(X,Y) = P(X)\cdot P(Y|X)$

$$\left.\begin{aligned} P(X,Y,E) &= P(X,Y|E)\cdot P(E) \\ P(X,Y,E) &= P(X|Y,E)\cdot P(Y|E)\cdot P(E) \end{aligned}\right\} \text{Product Rule}$$

Therefore : $P(X,Y|E)\cdot \cancel{P(E)} = P(X|Y,E)\cdot P(Y|E)\cdot \cancel{P(E)}$

$$\Rightarrow P(X,Y|E) = P(X|Y,E)\cdot P(Y|E)$$

(b) Prove $P(X|Y,E) = P(Y|X,E)P(X|E) / P(Y|E)$

Product Rule : $P(X,Y) = P(X)\cdot P(Y|X)$

$$\left.\begin{aligned} P(X,Y,E) &= P(X|Y,E)\cdot P(Y|E)\cdot P(E) \\ P(X,Y,E) &= P(Y|X,E)\cdot P(X|E)\cdot P(E) \end{aligned}\right\} \text{Product Rule}$$

$$\Rightarrow P(X|Y,E)\cdot P(Y|E)\cdot \cancel{P(E)} = P(Y|X,E)\cdot P(X|E)\cdot \cancel{P(E)}$$

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

$$\Rightarrow P(X|Y,E) = P(Y|X,E)\cdot P(X|E) / P(Y|E)$$

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

$$P(X,E) = P(X|E) \cdot P(E)$$

$$\sum_i P(X,E,Y=y_i) = P(X|E) \cdot P(E) \leftarrow \text{(marginalization rule)}$$

$$\sum_i P(X,Y=y_i|E) \cdot \cancel{P(E)} = P(X|E) \cdot \cancel{P(E)} \leftarrow \text{(product rule)}$$

$$\Rightarrow P(X|E) = \sum_i P(X, Y=y_i | E)$$

1.2 Prove ① ② ③ is equivalent :

(1) $P(X,Y|E) = P(X|E) P(Y|E)$ ①

Times $P(E)$ to Both side of equation ① :

$$\Rightarrow P(X,Y|E) \cdot P(E) = \underline{P(X|E) P(Y|E) \cdot P(E)} \; a = P(X,Y,E)$$

$$\Rightarrow P(X,Y,E) = \underline{P(X|Y,E) \cdot P(Y|E) \cdot P(E)} \; b$$

$$\Rightarrow P(X,Y,E) = \underline{P(Y|X,E) \cdot P(X|E) \cdot P(E)} \; c$$

Take $a = b$:

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

$$\Rightarrow P(X|Y,E) = P(X|E) \quad ②$$

Because ① $\Leftrightarrow$ $a=b$, $a=b \Leftrightarrow$ ②, ① $\Leftrightarrow$ ②

Take $a = c$:

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

$$\Rightarrow P(Y|X,E) = P(Y|E) \quad (3)$$

Because (1) $\Leftrightarrow$ $a=b=c$, $a=c \Rightarrow$ (3), (1) $\Rightarrow$ (2), (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3)

1.3

(a) $P(X=1) < P(X=1 \mid Y=1) < P(X=1 \mid Y=1, Z=1)$

X: Get to school on time or not, $X \in \{0,1\}$

Y: Bus on time or not, $Y \in \{0,1\}$

Z: previous good night sleep or not, $Z \in \{0,1\}$

The probability of getting to school on time knowing that the bus is on time and a previous good night sleep $(P(X=1 \mid Y=1, Z=1))$ is larger than the probability of getting to school on time only knowing that the bus is on time $(P(X=1 \mid Y=1))$, and larger than the probability of getting to school on time without additional info of bus and sleep quality $(P(X=1))$.

(b) $P(X=1 \mid Y=1) > P(X=1)$, $P(X=1 \mid Y=1, Z=1) < P(X=1 \mid Y=1)$

X: Get to school on time or not, $X \in \{0,1\}$
Y: Bus on time or not, $Y \in \{0,1\}$
Z: Previous night party $Z \in \{0,1\}$

The probability of going to school on time knowing that the bus is on time $(P(X=1 \mid Y=1))$ is larger than the probability of going to school on time without any additional information $(P(X=1))$, but the probability of going to school on time knowing that the bus is on time but the person spent last night partying $(P(X=1 \mid Y=1, Z=1))$ is smaller than the probability of going to school on time knowing that the bus is on time $(P(X=1 \mid Y=1))$.

(C) $P(X=1, Y=1) \neq P(X=1)(Y=1)$, $P(X=1, Y=1 \mid Z=1) = P(X=1 \mid Z=1) \cdot P(Y=1 \mid Z=1)$

X: Mountain fire going on or not, $X \in \{0,1\}$
Y: Dad was smoking or not, $Y \in \{0,1\}$
Z: Dad was at home or not, $Z \in \{0,1\}$

$P(X=1, Y=1) \neq P(X=1)P(Y=1)$, $P(X=1, Y=1 | Z=1) = P(X=1|Z=1) \cdot P(Y=1|Z=1)$

The probability of mountain fire going on knowing dad was smoking ($P(X=1, Y=1)$) is not equal to the probability of dad was smoking and mountain fire is going on just happened to be on the same day independently ($P(X=1) \cdot P(y=1)$) because it could be because dad smoked, and the sparkle of smoke increased the probability of mountain fire. However, the probability of mountain fire going on and dad smoked given that my dad was at home is the same as the probability of mountain fire going on while dad is at home happened to be the time after dad smoked at home $P(X=1|Z=1) \cdot P(Y=1|Z=1)$, since when at home, dad smoked and mountain fire is independent event.

continue →

1-4

(a)

$$\sum_i (T=i \mid D=d_j) = 1$$

T	$P(T \mid D=0)$
0	0.95
1	0.05

T	$P(T \mid D=1)$
0	0.1
1	0.9

$$\boxed{\begin{aligned} P(D=0) &= 0.99 \\ P(D=1) &= 0.01 \end{aligned}}$$

$$D \longrightarrow T$$

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

$= P(T=0 \mid D=0) \cdot P(D=0) + P(T=0 \mid D=1) \cdot P(D=1)$

$= 0.95 \cdot 0.99 + 0.1 \times 0.01 = 0.9415$

$\Rightarrow P(T=1) = 1 - 0.9415 = 0.0585$

$$P(D=0 \mid T=0) = \frac{P(T=0 \mid D=0) \cdot P(D=0)}{P(T=0)}$$

$= (0.95 \times 0.99) / 0.9415 = 0.9989 = 99.89\%$

(c) $$P(D=1 \mid T=1) = \frac{P(T=1 \mid D=1) \cdot P(D=1)}{P(T=1)}$$

$= (0.9 \cdot 0.01) / 0.0585 = 0.1538 = 15.38\%$

1.5

(a) $f(x) = -\sum_{i=1}^{n} p_i \log p_i$

$g(x) = \sum_i p_i = 1$

Lagrange multiplier

$L(p_i, \lambda) = -\sum_{i=1}^{n} p_i \log p_i - \lambda(\sum_i p_i - 1)$ [Lagrangian func]

① Take derivative against p_i:

$$0 = \frac{\partial L}{\partial p_i}$$

Since take derivative against p_i, all p_j ($j \in \{0 \dots n\}$ and $j \neq i$) becomes constant, therefore Σ sign goes away and we get:

$$0 = \frac{\partial}{\partial p_i}(-p_i \log p_i - \lambda(p_i - 1))$$

$$0 = -(1 + \log p_i) - \lambda \Rightarrow \log p_i = -1 - \lambda \Rightarrow \boxed{p_i = e^{-(1+\lambda)}}$$

② Take derivative against λ:

$$0 = \frac{\partial L}{\partial \lambda}$$

$$0 = \frac{\partial}{\partial \lambda}\left(-\sum_{i=1}^{n} p_i \log p_i - \lambda(\sum_i p_i - 1)\right)$$

$$0 = \frac{\partial}{\partial \lambda}(-\lambda(p_1 + p_2 + \dots p_n - 1)$$

$$0 = \frac{\partial}{\partial \lambda}(-\lambda p_1 - \lambda p_2 - \lambda n + \lambda) = -1(\sum_i^n p_i - 1)$$

$$\Rightarrow \boxed{\sum_i^n p_i = 1}$$

Given $\sum_i p_i = 1$ and $p_i = e^{-(1+\lambda)}$ [which is a constant]

when maximized $H[X]$, therefore $p_i = \frac{1}{n}$

5(b)

$$H(X_1, X_2 \cdots X_n) = -\sum_{X_1}\sum_{X_2}\cdots\sum_{X_n} P(X_1, X_2, \cdots X_n) \log P(X_1, X_2 \cdots X_n)$$

Given X_i independent from each other, then:

$$\log(P(X_1, X_2 \cdots X_n)) = \log(P(X_1) \cdot P(X_2) \cdots P(X_n))$$

$$\Rightarrow H(X_1, X_2 \cdots X_n) = -\sum_{X_1}\sum_{X_2}\sum_{X_n} P(X_1 X_2, \cdots X_n) \log(P(X_1) \cdot P(X_2) \cdots P(X_n))$$

$$\Rightarrow H(X_1, X_2 \cdots X_n) = -\sum_{X_1}\sum_{X_2}\sum_{X_n} P(X_1 X_2, \cdots X_n) \underbrace{(\log(P(X_1)) + \log(P(X_2)) + \cdots \log P(X_n))}_{\sum_{i=1}^{n} \log(P(X_i))}$$

$[\log(ab) = \log(a) + \log(b)]$

$$\Rightarrow H(X_1 X_2 \cdots X_n) = -\sum_{X_1}\sum_{X_2}\sum_{X_n} P(X_1 X_2 \cdots X_n) \log P(X_1) + P(X_1, X_2 \cdots X_n) \log P(X_2) + \cdots P(X_1, X_2 \cdots X_n) \log P(X_n)$$

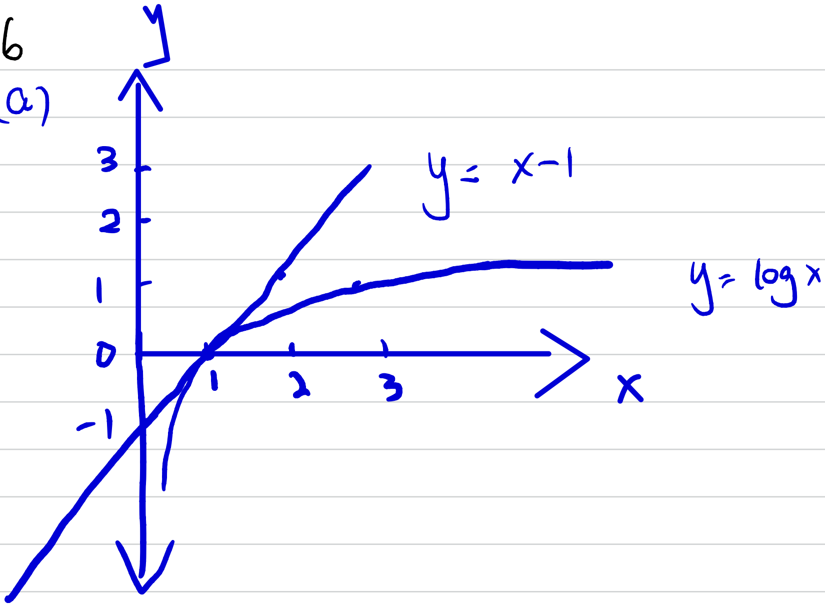
let $j \in \{1 \cdots n\}$, $j \neq i$, accumulate on X_j:

$$H(X_1 X_2 \cdots X_n) = -\sum_{i=1}^{n} (P(X_1) \log P(X_1) + P(X_2) \log P(X_2) + \cdots)$$

$$= \sum_{i=1}^{n} \left[-\sum_{X_i} P(X_i) \log P(X_i)\right] = \sum_{i=1}^{n} H(X_i)$$

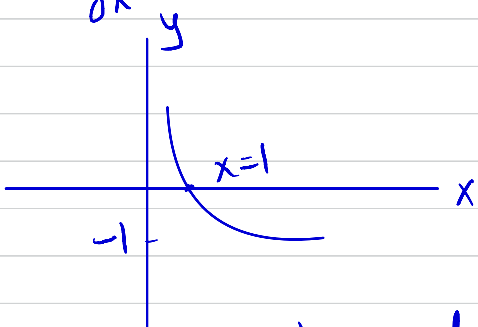
1.6

(a)



$y = \log(x) - (x-1)$

$\frac{\partial y}{\partial x} = \frac{1}{x} - 1$



[graph of derivative of $y = \log(x) - (x-1)$]

From the graph, the derivative is positive till $x = 1$, which means $\log(x) - (x-1)$ reaches maximum when $x = 1$

Therefore $\log(x) \leq x-1$ [$\log(x) = x-1$ when $x = 1$]

(b) Given: $\log(x) \le x-1$

$$KL(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$$

$$= \sum_i p_i (\log p_i - \log q_i)$$

$$= -\sum_i p_i (\log q_i - \log p_i)$$

$$= -\sum_i p_i \log\left(\frac{q_i}{p_i}\right)$$

$\because$ $\log(x) \le x-1 \Rightarrow -\log(x) \ge -(x-1)$

$$\therefore KL(p,q) = -\sum_i p_i \log\left(\frac{q_i}{p_i}\right) \ge -\sum_i p_i \left(\frac{q_i}{p_i} - 1\right) \quad a$$

$$\ge -\left(\sum_i q_i - \sum_i p_i\right)$$

$$\ge -1+1$$

$$\ge 0$$

if two distribution p_i and q_i is equal, $p_i = q_i$

$KL(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) = \sum_i p_i \log(1) = 0$

if $KL(p,q) = 0$, implied $KL(p,q) = -\sum_i p_i \left(\frac{q_i}{p_i} - 1\right)$ from a

Given $p_i \ge 0$ and $\sum p_i = 1$, $\left(\frac{q_i}{p_i} - 1\right) = 0$, which

implies $p_i = q_i$ given $KL(p,q) = 0$

As a result, $KL(p,q) = 0$ if and only if $p_i = q_i$

(C)

$$KL(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$$

$$-KL(p,q) = \sum_i p_i \log\frac{q_i}{p_i}$$

$$-KL(p,q) = 2\sum_i p_i \log\sqrt{\frac{q_i}{p_i}}$$ $\boxed{\text{By } \log x = 2\log\sqrt{x}}$

$$-KL(p,q) \leq 2\sum_i p_i \cdot \left(\frac{\sqrt{q_i}}{\sqrt{p_i}} - 1\right)$$ $\boxed{\text{Given } \log(x) \leq x-1}$

$$-KL(p,q) \leq \sum_i \left(2\sqrt{p_i}\sqrt{q_i} - 2p_i\right)$$

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

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

$$\geq \sum_i p_i - \sum_i 2\sqrt{p_i}\sqrt{q_i} + \sum_i q_i)$$

$$\geq \sum_i (\sqrt{p_i} - \sqrt{q_i})^2$$

d. $p_i = P(X=i \mid E)$, $q_i = P(X=i \mid E')$

$X \in (0,1)$

Let $p(0) = 0.2$ $p(1) = 0.8$

$q(0) = 0.5$ $q(1) = 0.5$

$$KL(p,q) = p_0 \log\left(\frac{p_0}{q_0}\right) + p_1 \log\left(\frac{p_1}{q_1}\right)$$

$$= 0.2 \cdot (-0.92) + 0.8 \cdot (0.47)$$

$$= -0.184 + 0.376$$

$$= 0.192$$

$$KL(q,p) = q_0 \log\left(\frac{q_0}{p_0}\right) + q_1 \log\left(\frac{q_1}{p_1}\right)$$

$$= 0.5 \cdot 0.92 + 0.5 \cdot (-0.47)$$

$$= 0.46 - 0.235$$

$$= 0.225$$

$\therefore KL(p q) \neq KL(q p)$

1.7

(a) $\sum_{x}\sum_{Y} P(x,y) = 1$ $\sum_{x}\sum_{Y} P(X)P(Y) = 1$

Let $p(x_i, y_i) = p_i$, $p(x_i)p(y_i) = q_i$

$I(X,Y) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$ $[0 < p_i, q_i \leq 1]$

Apply conclusion from 1.6 b, $I(X,Y) \geq 0$

Therefore $I(X,Y)$ is non-negative

(b) if X and Y independent: $P(X)\cdot P(Y) = P(X,Y)$

$I(X,Y) = \sum_{x}\sum_{y} P(x,y) \log\left[\frac{P(x,y)}{p(x)P(y)}\right]$

if $P(X)P(Y) = P(X,Y)$, $I(X,Y) = 0$ ✓

if: $I(X,Y) = 0$, it implies $P(X,y) - P(X)\cdot P(Y) = 0$

[from formula (a)], Given $P(x,y) > 0$ and $P(X)P(y) > 0$

$I(x,y) = 0$ only when $P(x,y) - P(x)P(y) = 0$, which $P(X,Y) = P(X)P(y)$ indicates independence ✓

Therefor $I(X,Y)$ vanishes if and only if X and Y are independent

1.8

(a) Yes, BN #1 implies Y and Z are conditional independent given X, which is not implied by BN#2

(b) No

(c) Yes, BN #3 implies X and Z are conditional independent given Y, which is not implied by BN #2

1.9

(a) The most frequent 15 words are :[('THREE', 273077), ('SEVEN', 178842), ('EIGHT', 165764), ('WOULD', 159875), ('ABOUT', 157448), ('THEIR', 145434), ('WHICH', 142146), ('AFTER', 110102), ('FIRST', 109957), ('FIFTY', 106869), ('OTHER', 106052), ('FORTY', 94951), ('YEARS', 88900), ('THERE', 86502), ('SIXTY', 73086)]

The least frequent 14 words are :[('BOSAK', 6), ('CAIXA', 6), ('MAPCO', 6), ('OTTIS', 6), ('TROUP', 6), ('CCAIR', 7), ('CLEFT', 7), ('FABRI', 7), ('FOAMY', 7), ('NIAID', 7), ('PAXON', 7), ('SERNA', 7), ('TOCOR', 7), ('YALOM', 7)] b.

correctly guessed	incorrectly guessed	best next guess ℓ	$P(L_i = \ell \text{ for some } i \in \{1,2,3,4,5\} \mid E)$
– – – – –	{}	E	0.5394
– – – – –	{E, A}	O	0.5340
A – – – S	{}	E	0.7715
A – – – S	{I}	E	0.7127
– – O – –	{A, E, M, N, T}	R	0.7454
– – – – –	{E, O}	I	0.6366
D – – I –	{}	A	0.8207
D – – I –	{A}	E	0.7521
– U – – –	{A, E, I, O, S}	Y	0.6270

c.

```python
# -*- coding: utf-8 -*-
"""
cs 250 hw1 1.9 c
"""

def count(all_counts):        # 1.9 a

    all_counts = sorted(all_counts, key=lambda x: x[1])
    least_frequent_15 = [item for item in all_counts[:14]]
    most_frequent_15 = [item for item in all_counts[len(all_counts)-15:]]
    most_frequent_15.reverse()
    return most_frequent_15, least_frequent_15
```

```python
def best_guess(words, correct_guessed, correct_guessed_pos, incorrect_guessed):   # 1.9 b

    words_satisfied_evidence = 0
    alphabet_match_count = [0] * 26

    for word in words:
        evidence = True
        for i, c in enumerate(correct_guessed):
            if word[0][correct_guessed_pos[i]] != c or word[0].count(c) > correct_guessed.count(c):
                evidence = False
                break

        if evidence:
            for c in incorrect_guessed:
                if c in word[0]:
                    evidence = False
                    break
        if evidence:

            words_satisfied_evidence += word[1]


            for c in set(word[0]):
                if c not in correct_guessed:
                    index = ord(c) - ord('A')
                    alphabet_match_count[index] += word[1]

    max_count = (0, None)
    for i, count in enumerate(alphabet_match_count):
        if count > max_count[0]:
            max_count = (count, chr(ord('A')+i))

    return max_count[1], round(max_count[0]/words_satisfied_evidence, 4)
```

```python
if __name__ == "__main__":
    file = open('hw1_word_counts_05-2.txt', 'r')
    all_counts = []
    for line in file:
        line.removesuffix('\n')
        word, c = line.split(' ')
        all_counts.append((word,int(c)))
    file.close()


    most_frequent, least_frequent = count(all_counts)
    print(f'The most frequent 15 words are :{most_frequent}\n\n\n',
          f'The least frequent 14 words are :{least_frequent}\n\n\n')


    best_guess_alphabet, probability = best_guess(all_counts, [], [], [])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')

    best_guess_alphabet, probability = best_guess(all_counts, [], [], ['E', 'A'])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')

    best_guess_alphabet, probability = best_guess(all_counts, ['A','S'], [0,4], [])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')

    best_guess_alphabet, probability = best_guess(all_counts, ['A','S'], [0,4], ['I'])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')

    best_guess_alphabet, probability = best_guess(all_counts, ['O'], [2], ['A', 'E', 'M', 'N', 'T'])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')

    best_guess_alphabet, probability = best_guess(all_counts, [], [], ['E', 'O'])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')


    best_guess_alphabet, probability = best_guess(all_counts, ['D','I'], [0, 3], [])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')


    best_guess_alphabet, probability = best_guess(all_counts, ['D','I'], [0, 3], ['A'])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')

    best_guess_alphabet, probability = best_guess(all_counts, ['U'], [1], ['A', 'E', 'I', 'O', 'S'])
    print(f'Best guess is {best_guess_alphabet}, the probablity is {probability}\n\n')
```