1.1)

a) To prove :- $P(X, Y|E) = P(X|E)P(Y|X, E)$

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

$$= \frac{P(X, E)\,P(Y|X, E)}{P(E)} \quad \boxed{P.R.}$$

$$= \frac{\cancel{P(E)}\,P(X|E)\,P(Y|X, E)}{\cancel{P(E)}} \quad \boxed{P.R.}$$

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

b) To prove :- $P(X|Y, E) = \dfrac{P(Y|X, E)\,P(X|E)}{P(Y|E)}$

$$P(X|Y, E) = \frac{P(X, Y|E)}{P(Y|E)} \quad \boxed{P.R.}$$

$$= \frac{P(X|E)\,P(Y|X, E)}{P(Y|E)} \quad \boxed{P.R.}$$

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

c) To prove :- $P(X|E) = \sum_y P(X, Y=y|E)$

$$P(X|E) = \frac{P(X,E)}{P(E)} \qquad \boxed{P.R.}$$

$$= \frac{\sum_y P(X, Y=y, E)}{P(E)} \qquad \boxed{Marg}$$

$$= \frac{\sum_y P(X, Y=y|E)\cancel{P(E)}}{\cancel{P(E)}} \qquad \boxed{P.R.}$$

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

1.2

(1) Given: $P(X,Y|E) = P(X|E)\,P(Y|E) \ldots (i)$

$$P(X,Y|E) = P(Y|E)\,P(X|Y,E) \quad \boxed{P.R.}$$

$\Rightarrow$ From (i)

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

$$\Rightarrow \boxed{P(X|Y,E) = P(X|E)} \quad \text{— (2)}$$

Furthermore;

$$P(X, Y|E) = P(X|E)\,P(Y|X, E) \quad \boxed{P.R.}$$

$\Rightarrow$ From ...(i)

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

$$\Rightarrow \boxed{P(Y|X, E) = P(Y|E)} \quad -③$$

② Given :- $P(X|Y, E) = P(X|E)$ ...(ii)

$$P(X|Y, E) = \frac{P(X, Y|E)}{P(Y|E)} \quad \boxed{P.R.}$$

From (ii)

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

$$\Rightarrow \boxed{P(X, Y|E) = P(X|E)\,P(Y|E)} \quad -①$$

Furthermore;

$$P(X \mid Y, E) = \frac{P(X, Y \mid E)}{P(Y \mid E)} \quad \boxed{P.R.}$$

$$= \frac{P(Y \mid X, E)\, P(X \mid E)}{P(Y \mid E)} \quad \boxed{P.R.}$$

From (ii)

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

$$\Rightarrow \boxed{P(Y \mid X, E) = P(Y \mid E)} \quad \text{— } (3)$$

(3) Given :- $P(Y \mid X, E) = P(Y \mid E)$ ... (iii)

$$P(Y \mid X, E) = \frac{P(X, Y \mid E)}{P(X \mid E)} \quad \boxed{P.R.}$$

From ... (iii)

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

$$\Rightarrow \boxed{P(X, Y|E) = P(X|E)P(Y|E)} \quad \text{— } 1$$

Furthermore;

$$P(Y|X,E) = \frac{P(X, Y|E)}{P(X|E)} \qquad \boxed{P.R.}$$

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

From (iii)

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

$$\Rightarrow \boxed{P(X|Y,E) = P(X|E)} \quad \text{— } 2$$

1.3

a) X = Car breaks down

Y = Car hasn't been serviced in a year

Z = Check engine light is on.

b) X = Car will break down

~~Z~~ Y = Clicking noise in the back

Z = Toy in the back.

c) X = Car breaks down

Y = Car gets in an accident

Z = AAA gets called.

1.4

a) $P(D=1) = 0.01$

$P(D=0) = 0.99$

D	$P(T=1 \mid D)$	$P(T=0 \mid D)$
0	0.05	0.95
1	0.90	0.10

$$D \longrightarrow T$$

b) $P(D=0 \mid T=0) = ?$

$$P(D \mid T) = \frac{P(T \mid D)\, P(D)}{P(T)} \quad \ldots \text{(i)} \quad \boxed{\text{Bayes}}$$

$$P(T) = \sum_{d} P(T, D=d) \qquad \boxed{\text{Marg}}$$

$$= \sum_{d} P(T \mid D=d)\, P(D) \quad \ldots \text{(ii)}$$

From i) & ii)

$$P(D|T) = \frac{P(T|D)\,P(D)}{\sum_{d} P(T|D=d)\,P(D=d)}$$

$$P(D=0|T=0) = \frac{P(T=0|D=0)\,P(D=0)}{P(T=0|D=0)\,P(D=0) + P(T=0|D=1)\,P(D=1)}$$

$$= \frac{0.99 \times 0.95}{0.95 \times 0.99 + 0.1 \times 0.01}$$

$$= \frac{0.99 \times 0.95}{0.9415}$$

$$= 0.9989$$

c) Similarily;

$$P(D=1|T=1) = \frac{P(T=1|D=1)\,P(D=1)}{P(T=1|D=0)\,P(D=0) + P(T=1|D=1)\,P(D=1)}$$

$$= \frac{0.9 \times 0.01}{0.05 \times 0.99 + 0.9 \times 0.01} = \frac{0.009}{0.0585} = \boxed{0.15}$$

$$\Downarrow$$

$$\boxed{0.1538}$$

5

1.5

$$H[X] = -\sum_i P_i \log P_i \quad \boxed{\text{function}}$$

$$\sum_i P_i = 1 \quad \boxed{\text{constraint}}$$

$$\Rightarrow f(x) = -\sum_i P_i \log P_i$$

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

$$\Rightarrow L(X) = f(x) - \lambda g(x)$$

$\Rightarrow$ To find the maxima :-

$$\nabla L = 0$$

$\Rightarrow$ For all the variables $P_1, P_2 \ldots P_n$ & λ, we'll differentiate

$\Rightarrow$

$$\frac{\partial \mathcal{L}}{\partial P_1} = 0 \Rightarrow -\log P_1 - 1 - \lambda = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial \mathcal{L}}{\partial P_n} = 0 \Rightarrow -\log P_n - 1 - \lambda = 0$$

$\Rightarrow$ For each P_i, we have

$$\log P_i = -(1+\lambda)$$

$$P_i = 2^{-(1+\lambda)} \quad \ldots (i)$$

Substituting (i) in the constraint

$$\sum_i P_i = 1$$

$$\Rightarrow \sum_i 2^{-(1+\lambda)} = 1$$

$$\Rightarrow 2^{-(1+\lambda)} \times n = 1 \Rightarrow n = 2^{(1+\lambda)} \Rightarrow \boxed{\lambda = \log_2 \frac{n}{2}} \ldots (ii)$$

Substituting (ii) in (i)

$$P_i = \frac{1}{2^{1+\log\frac{n}{2}}}$$

$$P_i = 2^{-\log_2 n}$$

$$\Rightarrow \log_2 P_i = -\log_2 n$$

$$\Rightarrow \boxed{P_i = \frac{1}{n}}$$

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

$\because X_1, X_2 \ldots X_n$ are independent

$$\Rightarrow -\sum_{X_1}\sum_{X_2}\ldots\sum_{X_n} P(X_1)P(X_2)\ldots P(X_n)\log\big(P(X_1)P(X_2)\ldots P(X_n)\big)$$

$$\Rightarrow -\sum_{X_1}\sum_{X_2}\ldots\sum_{X_n} P(X_1)P(X_2)\ldots P(X_n)\big(\log P(X_1) + \log P(X_2)\ldots \log P(X_n)\big)$$

$$\cancel{\Rightarrow -\sum_{X_1}\sum_{X_2}\ldots\sum_{X_n} P(X_1)P(X_2)\ldots P(X_n) - \sum_{X_1}\sum_{X_2}\ldots\sum_{X_n} P(X}$$

$$\Rightarrow -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} P(X_1)P(X_2)\cdots P(X_n)\log P(X_1)$$

$$-\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} P(X_1)P(X_2)\cdots P(X_n)\log P(X_2)$$

$$\vdots$$

$$-\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} P(X_1)P(X_2)\cdots P(X_n)\log P(X_n)$$

$$\Rightarrow -\sum_{x_1} P(X_1)\log P(X_1) - \sum_{x_2} P(X_2)\log(P(X_2)) - \cdots - \sum_{x_n} P(X_n)\log(P(X_n))$$

marg

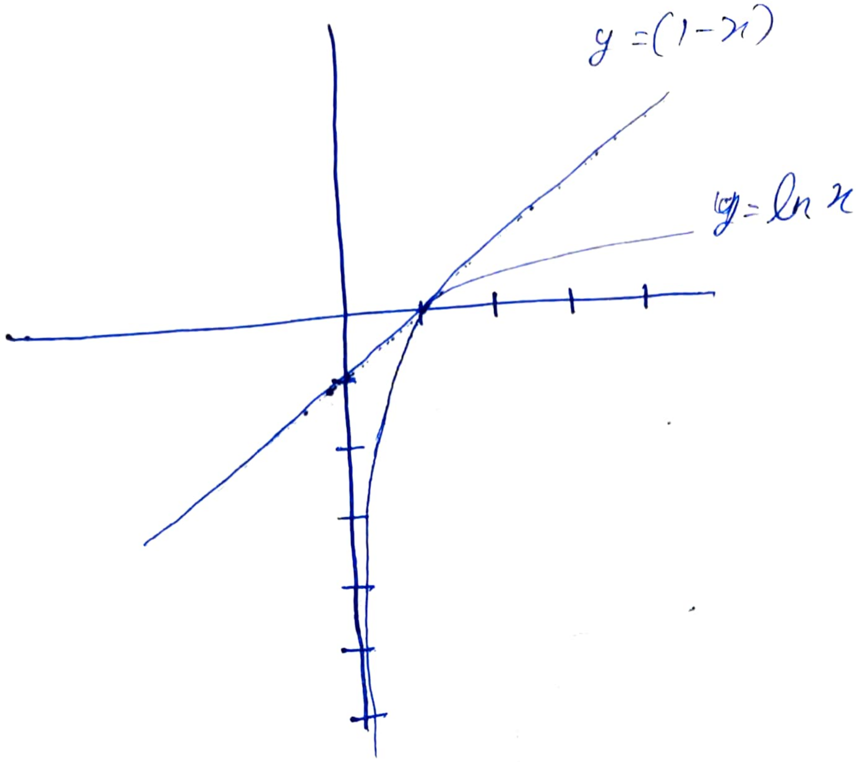
$$\Rightarrow H[X_1] + H[X_2] + \cdots H[X_n]$$

$$\Rightarrow \sum_i H[X_i]$$

$$\Rightarrow \boxed{H(X_1, X_2 \ldots X_n) = \sum_i H[X_i]}$$

1.6

a)



It can be seen that;

$$\log(x) \leq x - 1$$

& at $x = 1$

$$\log(x) = 0$$
$$x - 1 = 0$$

$\Rightarrow \log(x) = x - 1$ at $x = 0$

b) $KL(P,q) = \sum_i P_i \log(P_i/q_i)$

$\Rightarrow$ From (a)

$$\Rightarrow \sum_i P_i \log\left(\frac{P_i}{q_i}\right) \geq \sum_i P_i \left(1 - \frac{q_i}{P_i}\right)$$

$$\geq \sum_i P_i - \sum_i q_i$$

$$\boxed{KL(P,q) \geq 0}$$

c) $\log\sqrt{x} \leq \sqrt{x} - 1$ ... from (a)

$$\Rightarrow 2\log\sqrt{x} \leq 2(\sqrt{x} - 1)$$

$$\Rightarrow \log x \leq 2(\sqrt{x} - 1)$$

$$\Rightarrow \log\frac{1}{x} \geq 2(1 - \sqrt{x}) \quad \text{...(i)}$$

$$\Rightarrow \sum_i P_i \log\frac{P_i}{q_i} \geq 2\sum_i P_i \left(1 - \sqrt{\frac{q_i}{P_i}}\right)$$

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

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

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

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

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

Q) X = Falling while running

E = Running on road

E' = Running on sand

$P(X=1 | E=1) = 0.01 \quad P(X=0 | E=1) = 0.99$

$P(X=1 | E'=1) = 0.1 \quad P(X=0 | E=) = 0.9$

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

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

$$\Rightarrow KL(P,q) = 0.99 \times \log\left(\frac{0.99}{0.9}\right) + 0.01 \times \log\left(\frac{0.01}{0.1}\right)$$

$$KL(q,P) = 0.9 \times \log\left(\frac{0.9}{0.99}\right) + 0.1 \times \log\left(\frac{0.1}{0.01}\right)$$

$$\Rightarrow KL(P,q) = 0.1029$$

$$KL(q,P) = 0.2084$$

$$\Rightarrow \boxed{KL(P,q) \neq KL(q,P)}$$

1.7

a) $$I(X,Y) = \sum_x \sum_y P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

$$\geq \sum_x \sum_y P(x,y) \cdot \left(1 - \frac{P(x)P(y)}{P(x,y)}\right)$$

$$\geq \sum_x \sum_y P(x,y) - \sum_x \sum_y P(x)P(y)$$

$$\geq \sum_x \sum_y P(x,y) - \sum_x P(x) \sum_y P(y)$$

$$\geq 1 - 1 \times 1$$

$$\geq 1 - 1$$

$$\boxed{I(X,y) \geq 0}$$

b) if $I(X, Y) = 0$ ......(i)

$$\Rightarrow \sum_x \sum_y p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) = 0$$

$$\Rightarrow \sum_x \sum_y p(x)\,p(y|x) \log\left(\frac{\cancel{p(x)}\,p(y|x)}{\cancel{p(x)}\,p(y)}\right) = 0$$

$$\Rightarrow \sum_x p(x) \sum_y p(y|x) \log\left(\frac{p(y|x)}{p(y)}\right) = 0$$

$$\Rightarrow \sum_x p(x)\, KL(y|x, y) = 0$$

$\because$ $p(x)$ cannot be zero, from (i)

$$\Rightarrow KL(y|x, y) = 0$$

$$\Rightarrow \log\left(\frac{p(y|x)}{p(y)}\right) = 0$$

$$\Rightarrow p(y|x) = p(y)$$

$\Rightarrow$ $\boxed{\text{X and Y are independent.}}$

Again, if X and Y are independent.

$$\Rightarrow I(X,Y) = \sum_x \sum_y P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

$$= \sum_x \sum_y P(x,y) \log\left(\frac{\cancel{P(x)}\cancel{P(y)}}{\cancel{P(x)}\cancel{P(y)}}\right)$$

$$= \sum_x \sum_y P(x,y) \log 1$$

$$\boxed{I(X,Y) = 0}$$

1.8

Belief Network #1

From this network, the joint probability can be inferred as:-

$$P(X,Y,Z) = P(X)P(Y|X)P(Z|X) \quad ..(i)$$

From the general equation of joint probability, we can say that:-

$$P(X,Y,Z) = P(X)\,P(Y,Z|X)$$
$$= P(X)\,P(Y|X)\,P(Z|X,Y)$$

From (i).

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

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

$\Rightarrow$ Y and Z are conditionally independent on X.

Belief Network #2

From this network, joint probability can be inferred as:-

$$P(X,Y,Z) = P(X)\,P(Y|X)\,P(Z|Y) \quad \text{....(ii)}$$

From the general equation of joint probability we have:-

$$P(X,Y,Z) = P(X)\,P(Y,Z|X)$$
$$= P(X)\,P(Y|X)\,P(Z|X,Y)$$

$\Rightarrow$ From (ii)

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

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

$\Rightarrow$ Z and X are conditionally independent on Y

Belief Network #3.

From this network, joint probability can be inferred as:-

$$P(X,Y,Z)=P(Z)\,P(Y|Z)\,P(X|Y) \quad \ldots (iii)$$

From the general equation of joint probability

$$P(X,Y,Z)=P(Z)\,P(X,Y|Z)$$
$$=P(Z)\,P(Y|Z)\,P(X|Z,Y)$$

$\Rightarrow$ From (iii)

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

$\Rightarrow P(X|Y) = P(X|Z,Y)$

$\Rightarrow$ X and Z are conditionally independent on Y

Thus, for the question:-

a) Yes. First network establishes

$$P(Y,Z|X) = P(Y|X)\,P(Z|X)$$

while, second network estabishes.

$$P(X,Z|Y) = P(X|Y)\,P(Z|Y)$$

b) No. Both networks estabish:-

$$P(Y,Z|Y) = P(X|Y)\,P(Z|Y)$$

c) Yes. Similar to (a) different statements of conditional independence are established

In [ ]:

```
import pandas as pd
import string
```

In [ ]:

```
df = pd.read_csv('hw1_word_counts_05.txt', sep=' ', header=None)
```

In [ ]:

```
df['P_w'] = df[1] / (df[1].sum())
```

15 Most frequent 5-letter words

In [19]:

```
df.sort_values(1, ascending=False).reset_index()[[0, 1, 'P_w']].take([i for i in range(15)]
```

Out[19]:

	0	1	P_w
0	THREE	273077	0.035627
1	SEVEN	178842	0.023333
2	EIGHT	165764	0.021626
3	WOULD	159875	0.020858
4	ABOUT	157448	0.020542
5	THEIR	145434	0.018974
6	WHICH	142146	0.018545
7	AFTER	110102	0.014365
8	FIRST	109957	0.014346
9	FIFTY	106869	0.013943
10	OTHER	106052	0.013836
11	FORTY	94951	0.012388
12	YEARS	88900	0.011598
13	THERE	86502	0.011286
14	SIXTY	73086	0.009535

14 Least frequent 5-letter words

In [20]:

```
df.sort_values(1).reset_index()[[0, 1, 'P_w']].take([i for i in range(14)])
```

Out[20]:

	0	1	P_w
0	MAPCO	6	7.827935e-07
1	BOSAK	6	7.827935e-07
2	CAIXA	6	7.827935e-07
3	OTTIS	6	7.827935e-07
4	TROUP	6	7.827935e-07
5	CLEFT	7	9.132590e-07
6	FOAMY	7	9.132590e-07
7	CCAIR	7	9.132590e-07
8	SERNA	7	9.132590e-07
9	YALOM	7	9.132590e-07
10	TOCOR	7	9.132590e-07
11	NIAID	7	9.132590e-07
12	PAXON	7	9.132590e-07
13	FABRI	7	9.132590e-07

The lowest P(w) are associated with the least frequent words and vice-versa, which makes sense since least frequent words are least likely to be picked from a random sample and vice-versa.

In [21]:

```
def func(word, p_w, inclusive, excluding):
    for key in inclusive:
      for i in range(5):
        if key == word[i] and i not in inclusive[key]:
          return 0
        if i in inclusive[key] and word[i] != key:
          return 0

    for l in excluding:
      if l in word:
        return 0

    return p_w
```

In [22]:

```
def func2(word, P_W_E, remaining):
  alphabet_list = list(string.ascii_uppercase)

  result = []

  for alphabet in alphabet_list:
    occurs = False
    for pos in remaining:
      if alphabet == word[pos]:
        occurs = True
        result.append(P_W_E)
        break;

    if occurs == False:
      result.append(0)

  return result
```

In [23]:

```
def getNextBestGuess(df, inclusive, excluding):
  remaining = [i for i in range(5)]
  for key in inclusive:
    for pos in inclusive[key]:
      remaining.remove(pos)

  df['P_w'] = df[1] / (df[1].sum())
  df['P_W_E_num'] = df.apply(lambda x: func(x[0], x['P_w'], inclusive, excluding), axis=1)
  df['P_W_E'] = df['P_W_E_num'] / df['P_W_E_num'].sum()
  temp = df.apply(lambda x: func2(x[0], x['P_W_E'], remaining), axis=1, result_type='expanc
  for i, c in enumerate(string.ascii_uppercase):
    df[c] = temp[i]

  new_df = pd.DataFrame(columns = ['l', 'P_l_E'])
  for i, c in enumerate(string.ascii_uppercase):
    new_df = new_df.append({'l': c, 'P_l_E': round(df[c].sum(), 4)}, ignore_index=True)
  new_df = new_df.set_index('l')
  new_df.index.names = [None]
  index = new_df[new_df['P_l_E']!=1.0].idxmax()

  return (new_df.loc[index].index.tolist()[0], new_df.loc[index].values.tolist()[0])
```

In [24]:

```
processing_list = [[{}, []],
                   [{}, ['E', 'A']],
                   [{'A': [0], 'S': [4]}, []],
                   [{'A': [0], 'S': [4]}, ['I']],
                   [{'O': [2]}, ['A', 'E', 'M', 'N', 'T']],
                   [{}, ['E', 'O']],
                   [{'D': [0], 'I': [3]}, []],
                   [{'D': [0], 'I': [3]}, ['A']],
                   [{'U': [1]}, ['A', 'E', 'I', 'O', 'S']]]
```

In [25]:

```
print(("Next best guess l", "P(L_i)"))
for item in processing_list:
  print(getNextBestGuess(df, item[0], item[1]))
```

```
('Next best guess l', 'P(L_i)')
('E', [0.5394])
('O', [0.534])
('E', [0.7715])
('E', [0.7127])
('R', [0.7454])
('I', [0.6366])
('A', [0.8207])
('E', [0.7521])
('Y', [0.627])
```