1a.

V	P(A=1/V)	P(K=1/V)	P(L=1/V)	P(V)
0	1/3	1/2	1/3	1/2
1	2/3	5/6	1/3	1/2

1b.

P(V=1 / A=0,K=1, L=0) = $\frac{P(A=0,K=1,L=0\ /\ V=1)P(V=1)}{P(A=0,K=1,L=0)}$

= $\frac{P(A=0\ /\ V=1)P(K=1\ /\ V=1)P(L=0\ /\ V=1)P(V=1)}{P(A=0)P(K=1)P(L=0)}$

= $\frac{(1/3)(2/3)(5/6)(1/2)}{(1/2)(2/3)(2/3)}$

= **5/12**

P(V=0 / A=0,K=1, L=0) = $\frac{P(A=0,K=1,L=0\ /\ V=0)P(V=0)}{P(A=0,K=1,L=0)}$

= $\frac{P(A=0\ /\ V=0)P(K=1\ /\ V=0)P(L=0\ /\ V=0)P(V=0)}{P(A=0)P(K=1)P(L=0)}$

= $\frac{(2/3)(1/2)(2/3)(1/2)}{(1/2)(2/3)(2/3)}$

= **1/2**

We can say that message M does not carries a virus since P(V=1 / A,K,L) < P(V=0 / A,K,L)

1c.

P(V=1 / A=0,K=1, L=0) = 1/2

P(V=0 / A=0,K=1, L=0) = 1/2

We didn't get the same values as in problem B. we assumed that P(A,L,K / V) = P(A / V)P(L / V)P(K / V) which may not be true all the times. The samples are really low and we are missing a lot of combinations which tells us more information. The probability which we get from the table are assumptions from table.

1d.

P(V) = P(V=0) + P(V=1) = 1

The above constraint should be taken care. There are many constraints other than this, like the distributions should sum to 1, but for only the rest of the six values there aren't any.

1e.

The conclusion can be changed by only changing the value of A since all variables are independent. Changing A values will change the value of P(A / V), if we have A as 0 whenever V=1 then P(V=1 / A,K,L) > P(V=0 / A,K,L) and hence the answer of question 1b is changed.

1f.

To compute the probability distribution of P(A,L,K / V) we need the values of P(A/V), P(L / V) and P(K / V) for different values of A,L,K and V. For P(A/V) we have 4 values, for P(L / V) and P(K / V) has 4 values each. We need a minimum of 12 values to get the distribution of P(A,L,K / V).

1g.

In the independence assumption we are considering A, K, L to be independent. The conditional probability takes into account only one of the three variables A,L,K but V is dependent on A,L,K appearing together. P(A / V) or P(L / V) or P(K / V) is calculated based on the data which is gathered at different point of times. We want to know the an event {A,L,K} happening at the same time which will give P(A,L,K / V). In reality we need to consider an event which is happening at the same time but theoretically we are not considering the time into account, so the independent condition may not hold true in reality. The data could be one of the reason, if there is noise in the data collected this condition doesn't hold good.

2a.

H(W) = $\sum_{i=1}^{N} P(W = w_i) \log_N P(W = w_i)$
I am considering log to the base N for all the parts, where N is the size of Vocabulary.

The theoretical minimum value of H(W) is 0. If we have only one word in the English text article and the probability of rest of the words is zero.

The theoretical maximum value of H(W) is 1. If all the words in the vocabulary have equal probability to appear in English text article.

2b.

The maximum value of H(W) is 1. The article is A1 = {w1,w2,w3,w4,w5,w6}. Basically the article should contain all the words and every word is equally probable.

The minimum value of H(W) is 0. The article is A2 = {w1}. Basically the article should contain only one word and rest of the words probability should be zero.

2c.

Let us suppose we have only two words in our vocabulary V = {w1,w2}.

Let the two articles be A1 = {w1} and A2 = {w2}

The article A3 which is concatenation of A1 and A2 is A3 = {w1,w2}

For A3 the maximum value of Entropy is 1 in this case.

$$H(W) = - p(w1)\log(p(w1)) - p(w2)\log(p(w2))$$

$$= 1$$

3a.

The likelihood is given by the function

$$L(u) = \prod_{i=1}^{N} \frac{u^x e^{-u}}{x!}, \quad u > 0$$

Log likelihood is given by $l(u) = \sum_{i=1}^{N} (x_i * log(u)) \quad - N u \quad - \sum_{i=1}^{N} (log(x_i!))$

Derivative of log likelihood with respect to u is $\frac{d(l(u))}{du} = \frac{\sum_{i=1}^{N} x_i}{u} - N$

and equating derivative to zero we have

$$u = \frac{\sum_{i=1}^{N} x_i}{N}$$

3b.

The likelihood is given by the function

$$L(u) = \prod_{i=1}^{N} \frac{u^x e^{-u}}{x!} (\lambda e^{-\lambda u}), \quad u > 0$$

Log likelihood is given by $l(u) = \sum_{i=1}^{N} (x_i * log(u)) \quad - N u \quad - \sum_{i=1}^{N} (log(x_i!)) - \sum_{i=1}^{N} (\lambda u \log(\lambda))$

Derivative of log likelihood with respect to u is $\frac{d(l(u))}{du} = \frac{\sum_{i=1}^{N} x_i}{u} - N - N\lambda\log(\lambda)$

and equating derivative to zero we have

$$u = \frac{\sum_{i=1}^{N} x_i}{N\,(1+\lambda\log(\lambda))}$$