Problem 1 Expected Length of Coding Scheme

(a) (a) Suppose that the symbols occur with probabilities Pr[A] =0.4 , Pr[B] = 0.2, Pr[C] = 0.2, Pr[D] = 0.1, and Pr[E] = 0.1, and the coding scheme encodes these symbols into binary codes as follows.

$$E[X_i] = 3(0.4) + 3(0.2) + 3(0.2) + 3(0.1) + 3(0.1) = 3$$

Then then value for $E[X_i]$ is 3. Therefore, using the fact that

$$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

Then it follows that

$$E[X] = \sum_{i=1}^{n} 3 = 3\sum_{i=1}^{n} 1 = 3n$$

Therefore, the expected number of bits required is 3n.

(b) Now suppose that the symbols occur with the same probabilities Pr[A] =0.4 , Pr[B] = 0.2, Pr[C] = 0.2, Pr[D] = 0.1, and Pr[E] = 0.1, but we have a different encoding scheme:
From the given then

$$E[X_i] = 1(0.4) + 0.2(2) + 0.2(3) + 4(0.1) + 4(0.1) = 3.4$$

Then the value for $E[X_i]$ is 3.4.Therefore, using the fact from part (a) then

$$E[X] = \sum_{i=1}^{n} 3.4 = 3.4\sum_{i=1}^{n} 1 = 3.4n$$

Thus, the expected number of bits required is 3.4n.

(c) Now consider a different information system that generates symbols with probabilities Pr[A] =0.5 , Pr[B] = 0.3, Pr[C] = 0.1, Pr[D] = 0.05, and Pr[E] = 0.05.
From the given coding scheme then:

$$E[X_i] = 1(0.5) + 2(0.3) + 3(0.1) + 4(0.05) + 4(0.05) = 1.8$$

Then, $E[X_i] = 1.8$ then from the formula used in part (a) then

$$E[X] = \sum_{i=1}^{n} 1.8 = 1.8\sum_{i=1}^{n} 1 = 1.8n$$

The expected number of bits required is 1.8n.

Problem 2 Random Gene Sequence

From the given there are 4 parts of the DNA and the probability to get A at first is

$$\Pr[A] = 0.25$$

and the probablity of the other letters are

$$\Pr[T, G, C] = 0.75$$

Then, we consider three cases.

- Case 1: we get letters besides A then that increases the expected value y E+1.
- Case 2: Other letter appears on the second pick therefore increasing it by $E + 2$
- Case 3: A appears twice in a row then the probability will be $(0.25)^2 = 0.0625$

Therefore we have the following equation of

$$E = 0.75(E + 1) + 0.75^2(E + 2) + 0.0625(2)$$

then algebraically solving for E results to

$$E = 20$$

Therefore, the expected value of getting two As in a row is E = 20.

Problem 3 Hashing with Chaining

(a) Consider a hash table with m slots that uses chaining for collision resolution. The table is initially empty. What is the probability that, after k keys are inserted, there is a chain of size k? Include an argument for or proof of your solution.
The probability is $\frac{1}{m^{k-1}}$ because there are m slots and let i be the location where the k key is slotted in.

(b) For h(k) = k mod 11 such that $k = \{20, 51, 10, 19, 32, 1, 66, 40\}$

- h(20) = 20 mod 11 = 9
- h(51) = 51 mod 11 = 7
- h(10) = 10 mod 11 = 10
- h(32) = 32 mod 11 = 10
- h(1) = 1 mod 11 = 1
- h(66) = 66 mod 11 = 0
- h(40) = 40 mod 11 = 7

h(k)	Linked List Cells			
0	66			
1	1			
2				
3				
4				
5				
6				
7	40	51		
8	19			
9	20			
10	32	10		

Problem 4 Open Address Strategies

(a) Show the table that results when 20, 51, 10, 19, 32, 1, 66, 40 are cumulatively inserted in that order into an initially empty hash table of size 11 with linear probing

32	1	68	40				51	19	20	10
0	1	2	3	4	5	6	7	8	9	10

(b) How many re-hashes after collision are required for this set of keys?

- h(20) = 20 mod 11 = 9
- h(51) = 51 mod 11 = 7
- h(10) = 10 mod 11 = 10
- h(19) = 19 mod 11 = 8
- h(32) = 32 mod 11 = 10 rehash
- h'(32,1) = (32+1) mod 11 = 33 mod 11 = 0
- h(1) = 1 mod 11 = 1
- h(66) = 66 mod 11 = 0 rehash
- h'(66,1) = (66+1) mod 11 = 67 mod 11 = 1 rehash
- h'(66,2) = (66+2) mod 11 = 68 mod 11 = 2
- h(40) = 40 mod 11 = 7 rehash
- h'(40,1) = (40+1) mod 11 = 41 mod 11 = 8 rehash
- h'(40,2) = (40 + 2) mod 11 = 42 mod 11 = 9 rehash
- h'(40,3) = (40+3) mod 11 = 43 mod 11 = 10 rehash
- h'(40,4) = (40+ 4) + 44 mod 11 = 0 rehash
- h'(40,5) = (40+5) mod 11 = 45 mod 11 = 1 rehash
- h'(40,6) = (40+6) mod 11 = 46 mod 11 = 2 rehash
- h'(40,7) = (40+7) mod 11 = 47 mod 11 = 3

Total Amount of Rehash 10 times

(c) Show the table that results when 20, 51, 10, 19, 32, 1, 66, 40 are cumulatively inserted in that order into an initially empty hash table of size m = 11 with double hashing and

66		1	40	32			51	19	20	10
0	1	2	3	4	5	6	7	8	9	10

(d) How many re-hashes after collision are required for this set of keys?

- h(20,0) = (20 mod 11 + 0(1 + (20 mod 7)) mod 11 = 9

- h(51,0) = (51 mod 11 + 0(1 + (51 mod 7)) mod 11 = 7
- h(10,0) = (10 mod 11 + 0(1 + (10 mod 7)) mod 11 = 10
- h(19,0) = (19 mod 11 + 0(1 + (19 mod 7)) mod 11 = 8
- h(32,0) = (32 mod 11 + 0(1 + (32 mod 7)) mod 11 = 10 rehash
- h(32,1) = (32 mod 11 + 1(1 + (32 mod 7)) mod 11 = 4
- h(1,0) = (1 mod 11 + 0(1 + (1 mod 7)) mod 11 = 1
- h(66,0) = (66 mod 11 + 0(1 + (66 mod 7)) mod 11 = 0
- h(40,0) = (40 mod 11 + 0(1 + (40 mod 7)) mod 11 = 7 rehash
- h(40,1) = (40 mod 11 + 1(1 + (40 mod 7)) mod 11 = 2

Total Amount of Rehash 2 times

(e) Recall from lecture that for an $\alpha = \frac{n}{m}$ then we have that the theoretical is

$$\frac{1}{1-\alpha}$$

Since when 40 is inserted there are 7 slots already inserted in m = 11 then

$$\frac{1}{1-\frac{7}{11}} = 2.75$$

Therefore, the expected number is 2.75.