

# Assignment 1

Mytraya Gattu, 180050032

22 August 2020

All files pertaining to the assignment can be found at <https://github.com/LordThunder333/PH549-assignment1.git>

## 1 Problem 1

In what follows,  $r$  is the radius of the hemispherical caps,  $l$  the length and  $m$  of the cell. It is given that,

$$r = 0.5\mu\text{m}$$

$$l = 2\mu\text{m}$$

(a) Volume is given by

$$V = \pi r^2 l + \frac{4}{3} \pi r^3 = 2.094\mu\text{m}^3$$

Typical cell volume is  $0.6 - 0.7\mu\text{m}^3$

[https://en.wikipedia.org/wiki/Escherichia\\_coli](https://en.wikipedia.org/wiki/Escherichia_coli)

(b)  $m$  is the sum of the mass of water contained and that of the other substances. Since, density of water is

$$10^3 \frac{\text{kg}}{\text{m}^3} = 10^{-6} \frac{\text{pg}}{\mu\text{m}^3}$$

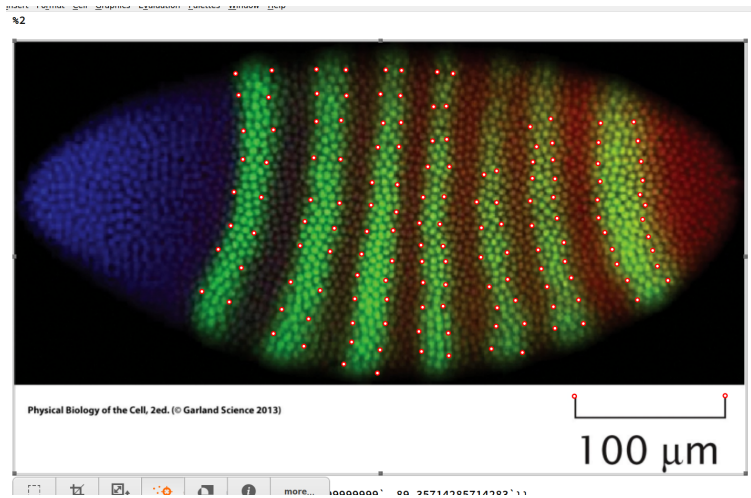
We have

$$m = 10^{-6} \cdot \left( \frac{2}{3} + 1.3 \cdot \frac{1}{3} \right) \cdot V_{\text{pg}} = 2.304 \cdot 10^{-3} \text{pg}$$

Typical cell weight is  $1 \cdot 10^{-3} \text{pg}$

[https://ecmdb.ca/e\\_coli\\_stats](https://ecmdb.ca/e_coli_stats)

It is evident from the above calculations, that the it is the assumed shape of *E. coli* which deviates greatly from what is true.



## 2 Problem 2

Let at the end of 9<sup>th</sup> cycle there be  $N$  nuclei at the surface. At the end of 13<sup>th</sup> cycle, therefore there are  $16N$  nuclei, which is given as  $\approx 6000$ . Hence,

$$N \approx 375$$

Since, every embryo must start from a single fertilized nucleus, at the end of 9<sup>th</sup> cycle, there must be  $2^9 = 512$  nuclei in total. Therefore, the fraction of nuclei which migrated to the surface is

$$\frac{375}{512} \approx 0.7324$$

Following the notation of Problem 1, surface area of the spherocylinder is

$$A = 4\pi r^2 + 2\pi r l = 4.3982 \cdot 10^{-7} \text{m}^2$$

Therefore, the areal density is

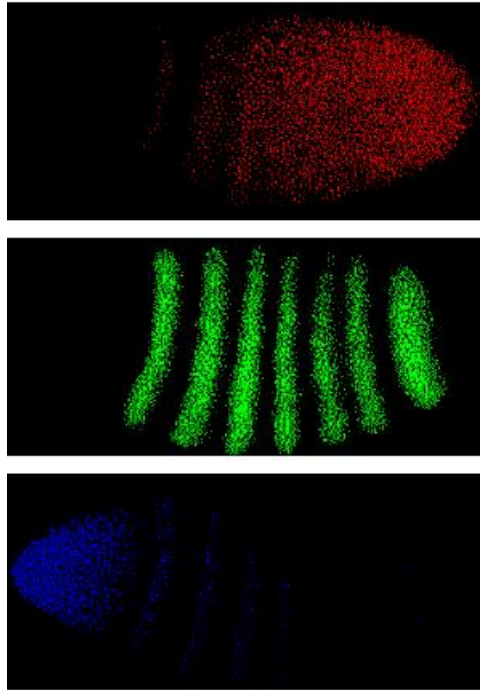
$$1.3642 \cdot 10^{10} \frac{\text{nuclei}}{\text{m}^2}$$

(For the next part, I have used Mathematica.) To calculate the length of the green strips, I marked the ends of widths along the lengths of each strip, and then used the given scale.

Length of the green strip is:  $18.37 \pm 4.99 \mu\text{m}$  To evaluate the number of nuclei in each strip, I isolated the colors, and counted the number of pixels.

Therefore, the number of nuclei corresponding to each dye(I'm considering only a single surface – In the diagram, the surface of the cylinder has been mapped to a rectangle)

- Red: 466
- Green: 660
- Blue: 182



### 3 Problem 3

#### 3.1 *E. coli*

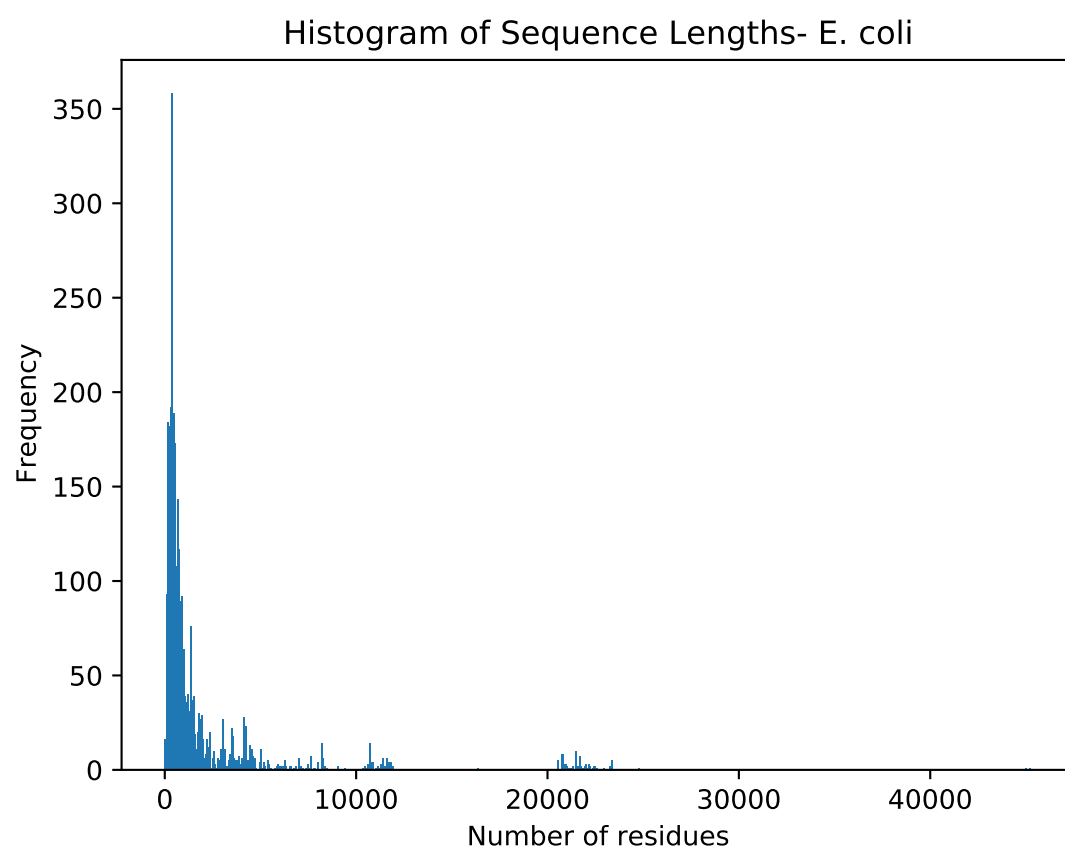
Mean of sequence length: 1892.00  
Standard deviation of sequence length: 3872.73  
Mean of molecular mass: 289.03  
Standard deviation of molecular mass: 775.76

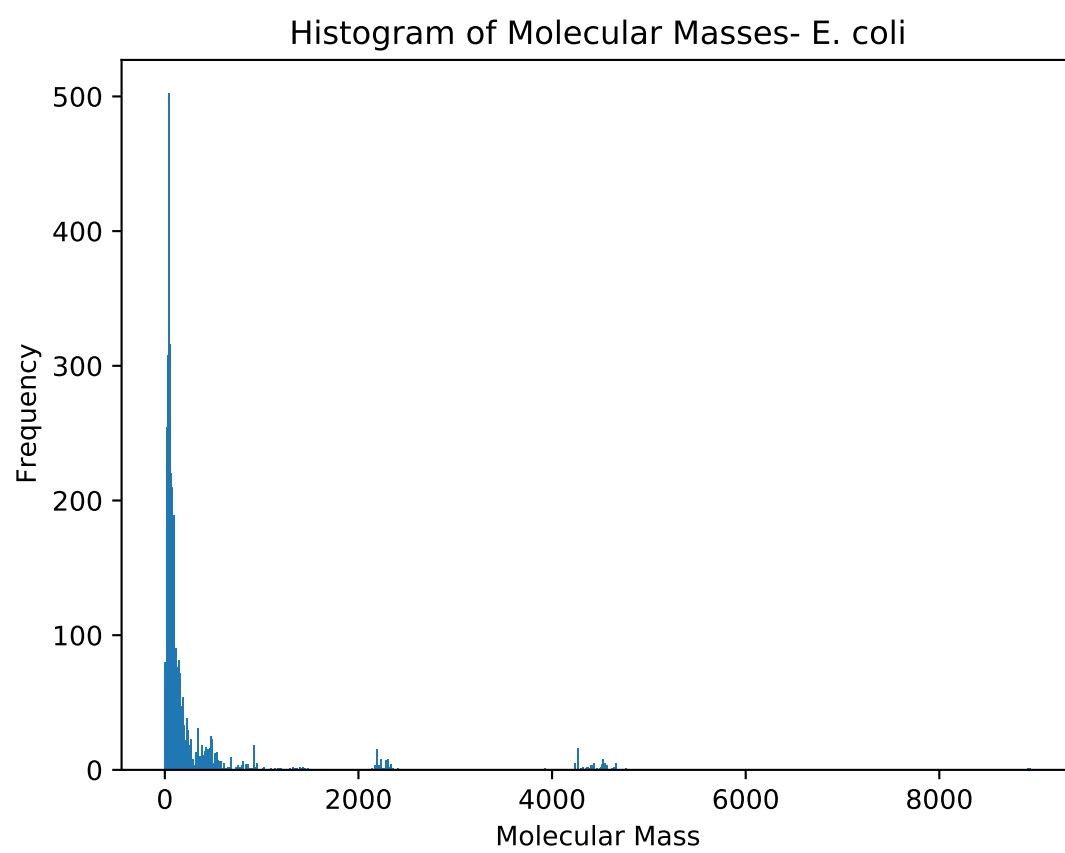
#### 3.2 *S. cerevisiae*

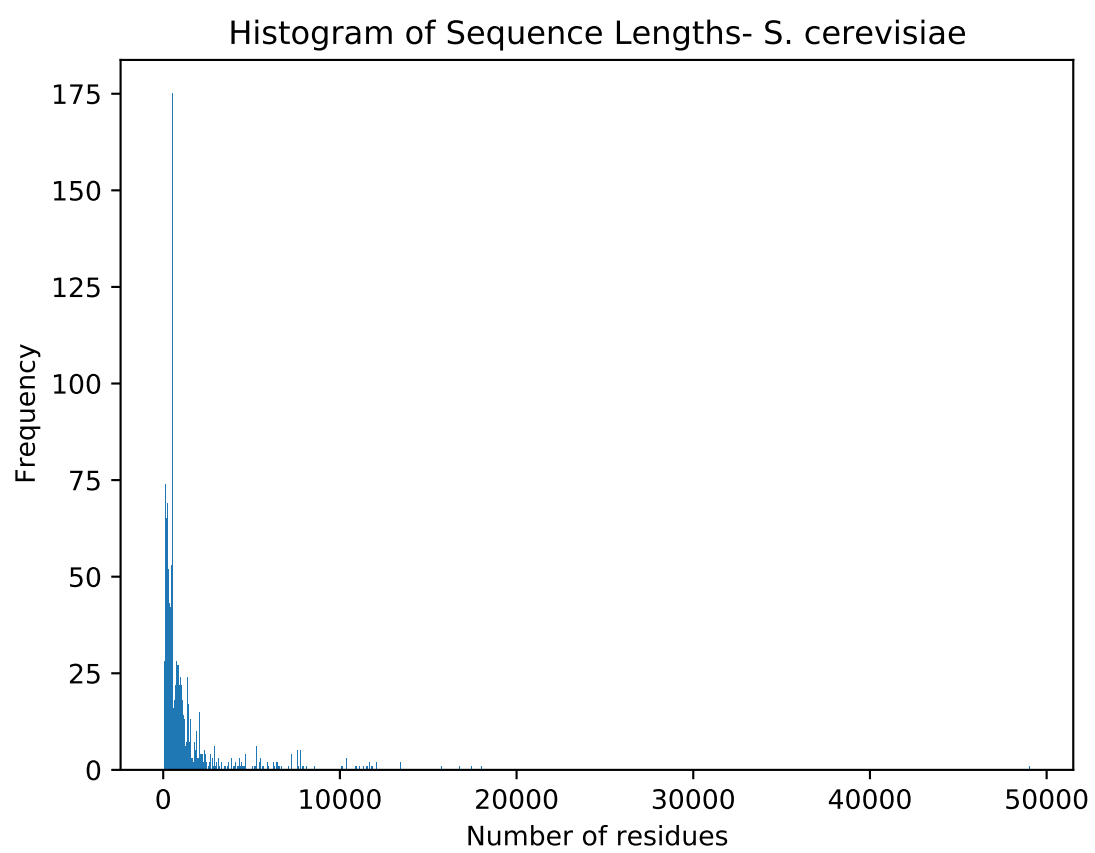
Mean of sequence length: 1287.90  
Standard deviation of sequence length: 2444.49  
Mean of molecular mass: 154.78  
Standard deviation of molecular mass: 313.31

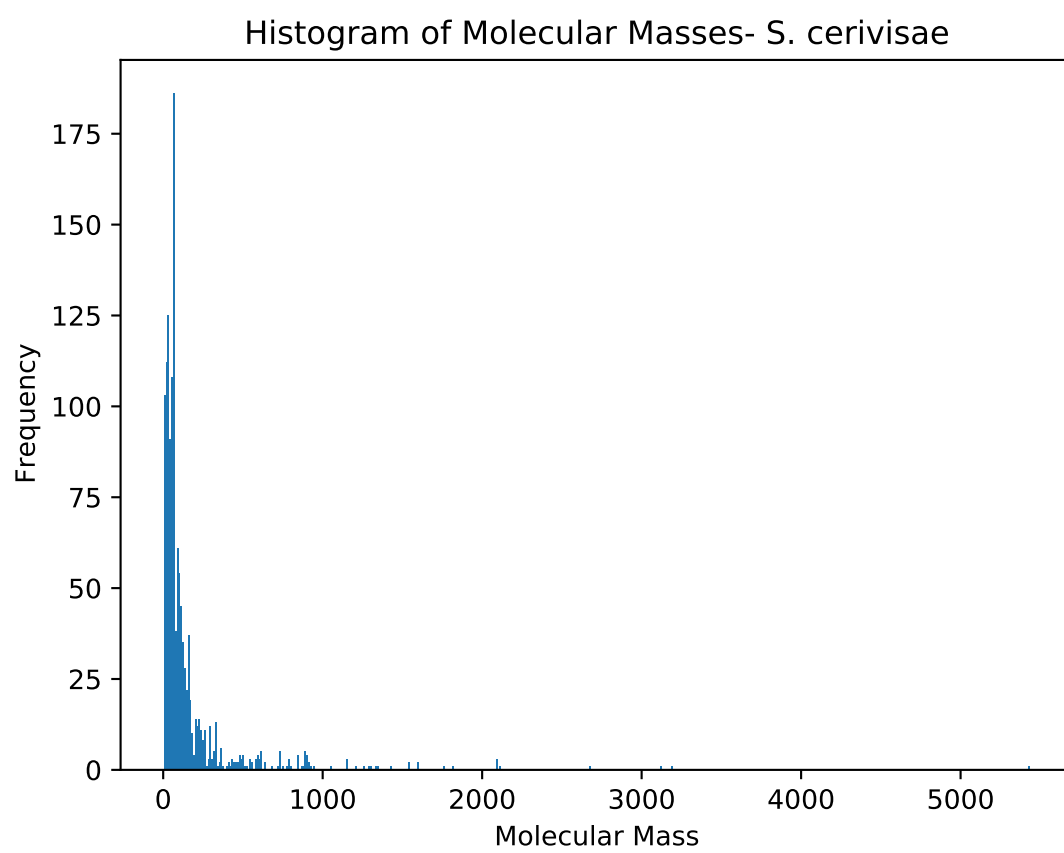
#### 3.3 Humans

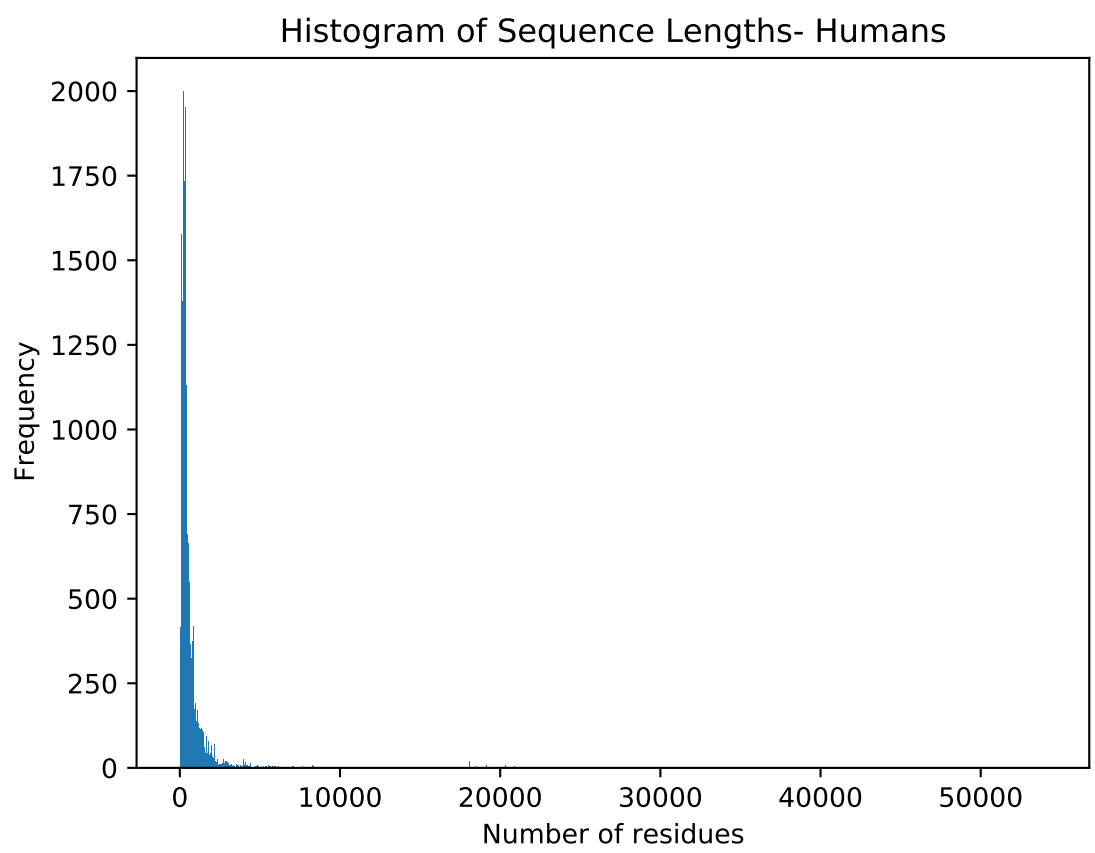
Mean of sequence length: 630.74  
Standard deviation of sequence length: 1263.08  
Mean of molecular mass: 74.21  
Standard deviation of molecular mass: 181.93



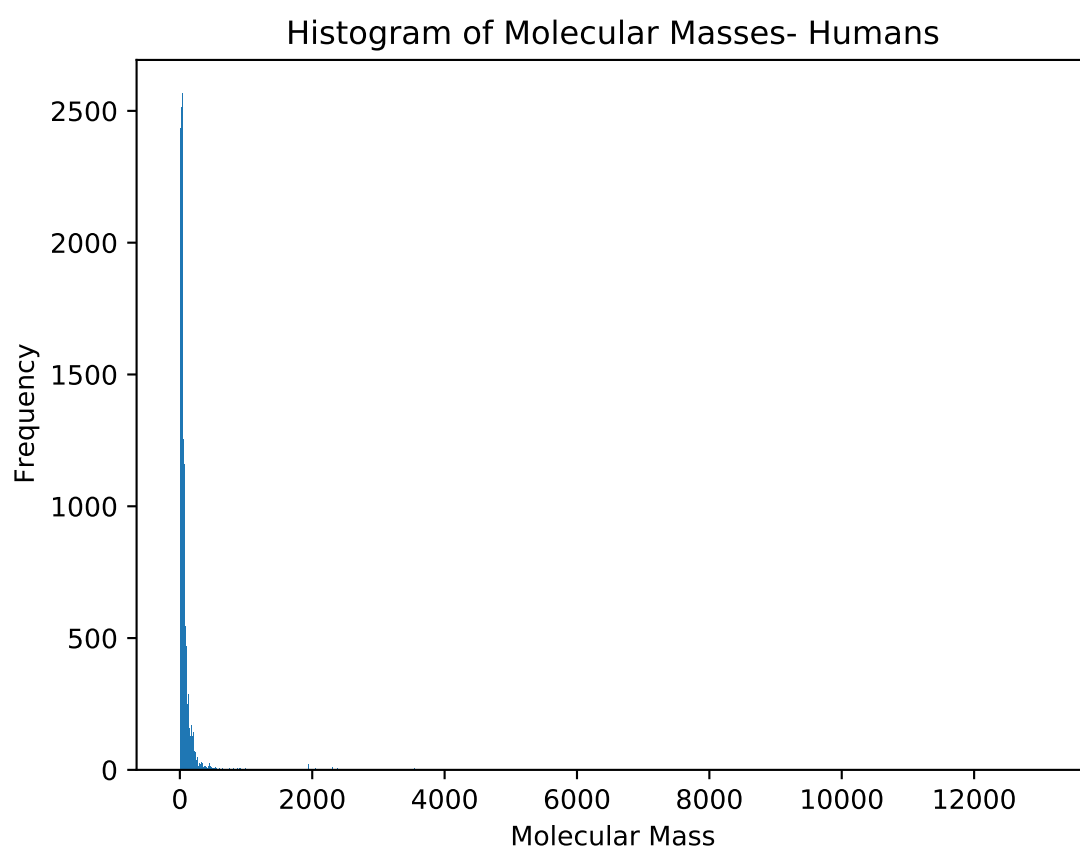












## 4 Problem 4

The probability of having 0 and 6 is

$$p^3 q^3$$

. The probability of having 1 and 5 is

$$\binom{3}{1} p^4 q^2 + \binom{3}{1} p^2 q^4$$

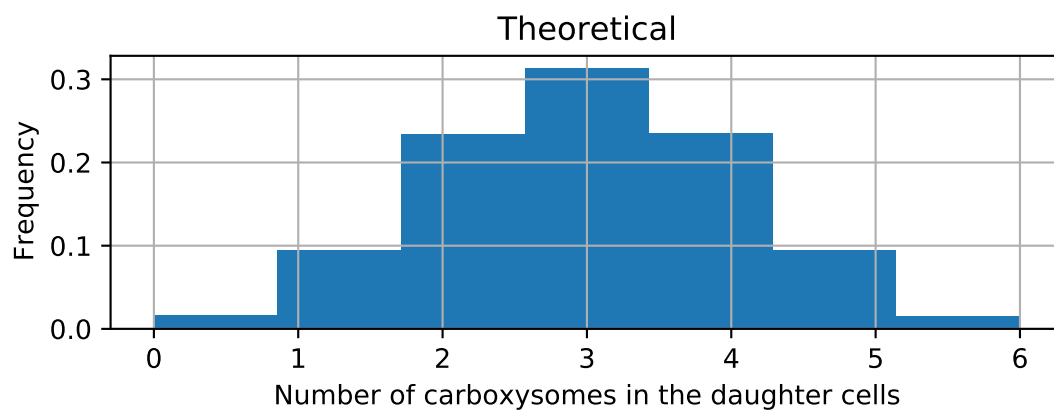
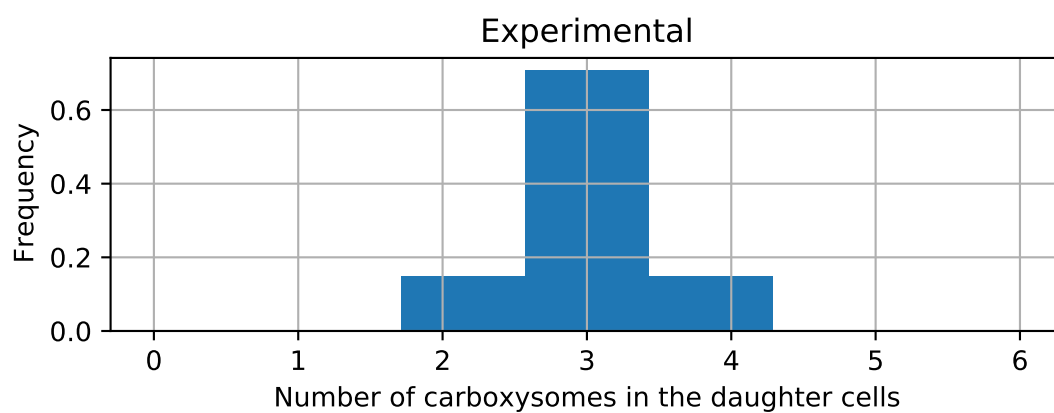
The probability of having 2 and 4 in one cell is

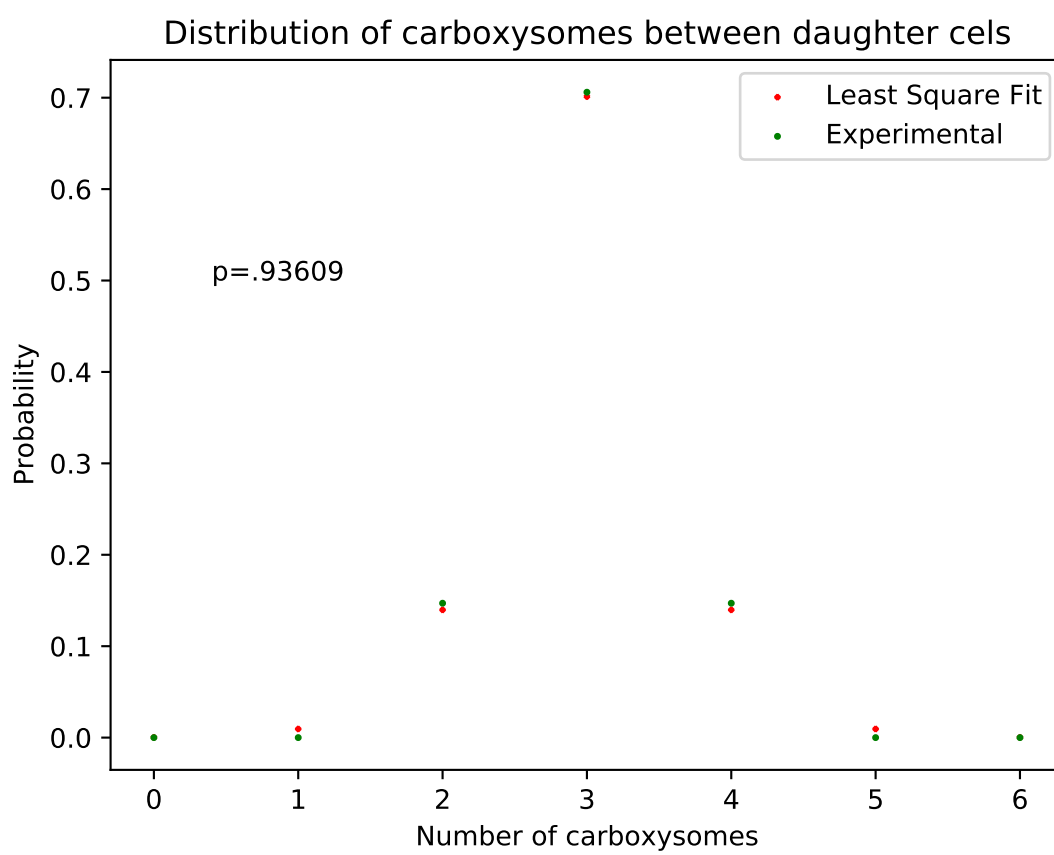
$$\binom{3}{2} p^5 q^1 + \binom{3}{2} p^1 q^5 + \binom{3}{1} \binom{3}{1} p^3 q^3$$

The probability of having 3 in one cell is

$$p^6 + q^6 + \binom{3}{1} \binom{3}{2} p^2 q^4 + \binom{3}{2} \binom{3}{1} p^4 q^2$$

By substituting  $p = 0.5 + x$ , and minimizing using the least-squares method with respect to the experimental distribution, obtained value for p is 0.936093. Attached, is a scatter plot showing the deviation. Given that, there are only 6 molecules, and the error is about  $0.007 \pm 0.001$  (for the non-zero values), I believe it's a good model.





## 5 Problem 5

- (a) There are  $4^3$  possible ways of choosing the bases to form a triplet, each with the same probability. Therefore,

$$p_s = \frac{3}{64}$$

- (b) There are 64 possible codons. Three of these are not allowed (The stop codons), to form an ORF. Therefore, the number of acceptable possibilities is  $(64 - 3)^N$ .

Required probability is  $\left(\frac{61}{64}\right)^N$ .

- (c) Codons are of length 3 bases. Since, the DNA is circular, let us pick an arbitrary base and call it the starting point, and label the bases as

$$\dots x_{N-1} x_N x_0 x_1 x_2 \dots$$

Looking at the above sequence, we can see that the codons containing  $x_0$  are

$$x_{N-1} x_N x_0$$

$$x_N x_0 x_1$$

$$x_0 x_1 x_2$$

and their reverse (corresponding to either a clockwise read or an anti-clockwise read – Also respectively, each codon corresponds to reading frames: +2, +1 and +0). Therefore, there are 6 reading frames in total.

- (d) Exact length of the E. coli genome is 4639675 (As per the data in U00096.fna) base pairs, whose remainder with respect to 3 is 1.

Since, when chosen at random, choice of reading frame doesn't matter, let's suppose that I have picked one already. In any reading frame, because of the non-zero remainder, when codons are selected, one base pair will remain ungrouped. I select this as my origin and group bases accordingly. The sequence will look like

$$\dots |XXX|X|XXX|\dots$$

Total number of codons formed is 1546558. Since, this is a very large number, the distribution to a good approximation must be poisson or a sum of poisson distributions. That is, for an ORF of length  $l$ , the probability that there will be  $N$  such ORFs, when a sequence is chosen at random must be

$$\sum_i e^{-\lambda_i(l)} \lambda_i^N(l) / N!$$

- (e) Take a look at the attached ipython notebook.  
 (f) The figures are placed at the end.  
 (g) 1  
 (h) Plot is placed at the end.

