

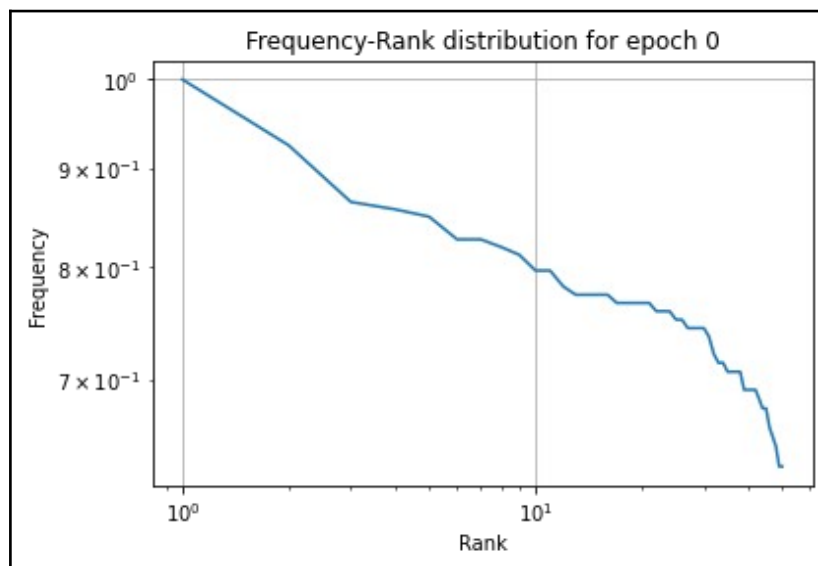
Analysis

Question 1:

- Total Number of recipes = 39774
- Epochs = 6
- Number of recipes per epoch = 6629
- Number of unique ingredients given = 6714
- Preprocessing done: The dataset was such that Onions and onion were treated separately, so **Lemmatization and Lower case conversion** was implemented which eventually reduced the Nature Basket size to **6703**
- Recipe Size = 10

Generation of Initial Primordial Cuisine:

- Fitness value to each ingredient of Nature Basket is assigned uniformly randomly.
- Primordial Cuisine size = 500
- Kitchen Basket size = 50
- The frequency Rank Distribution for initial primordial cuisine is as follows:

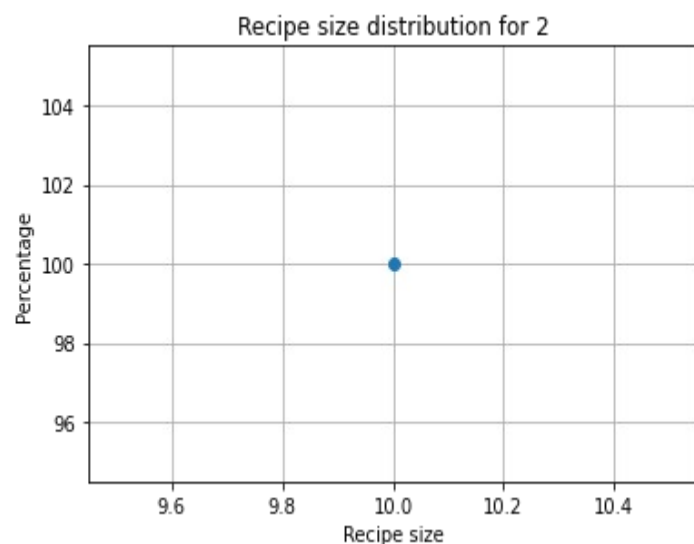
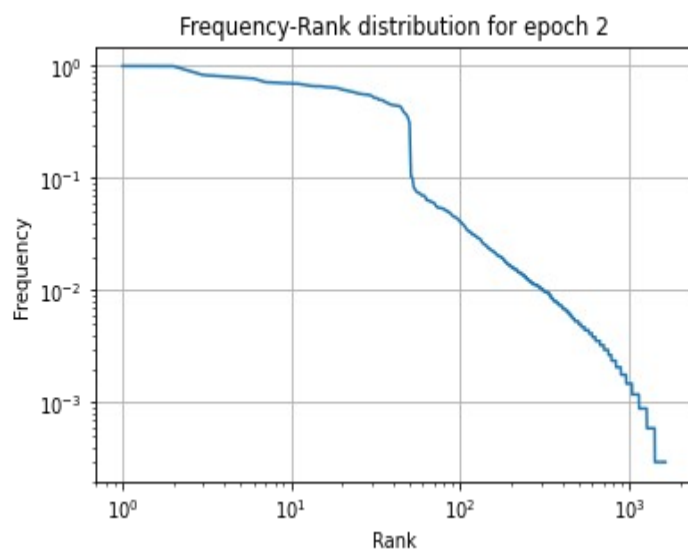
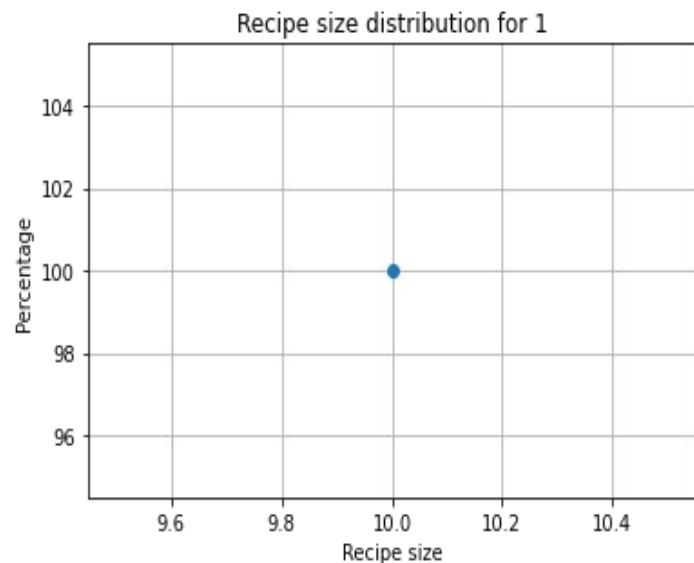
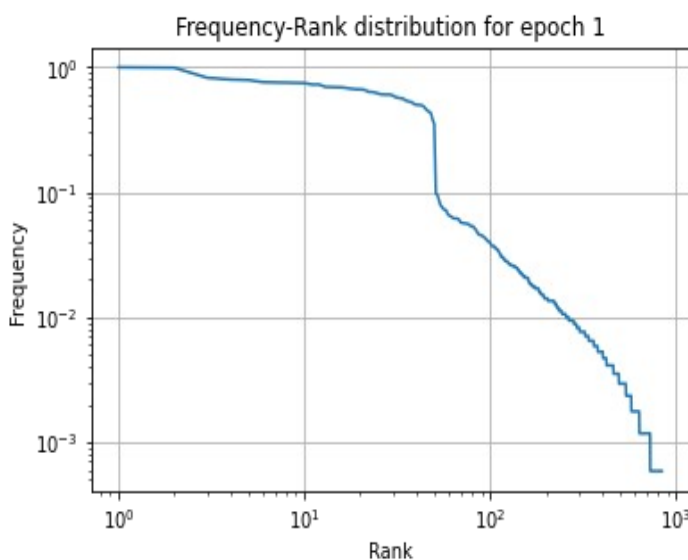


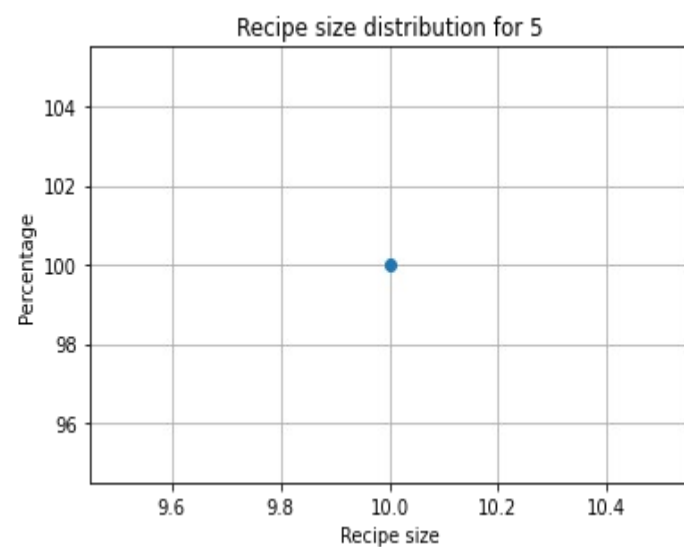
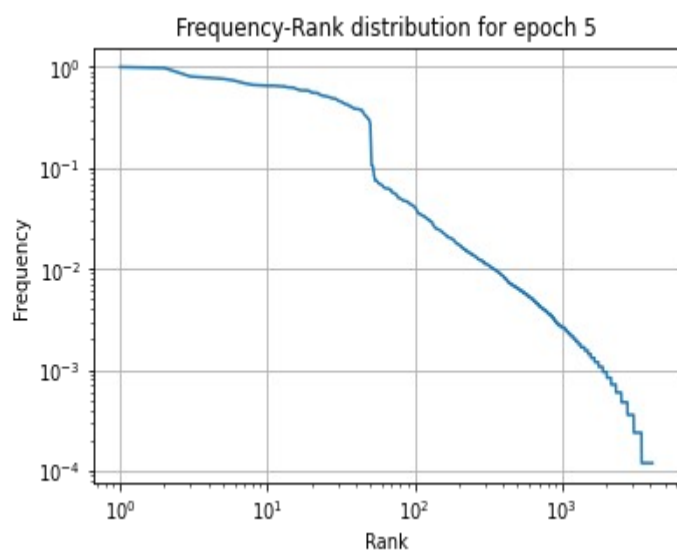
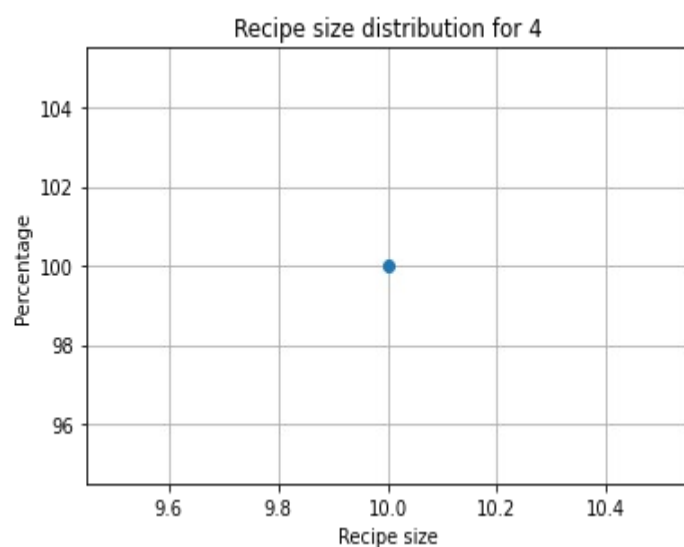
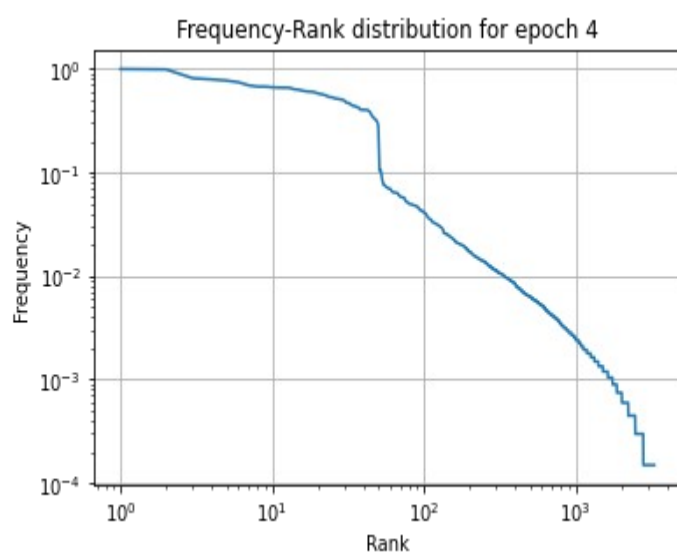
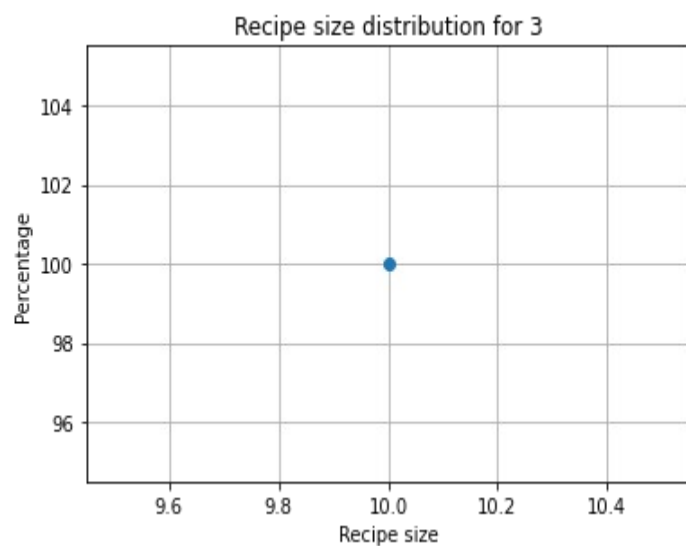
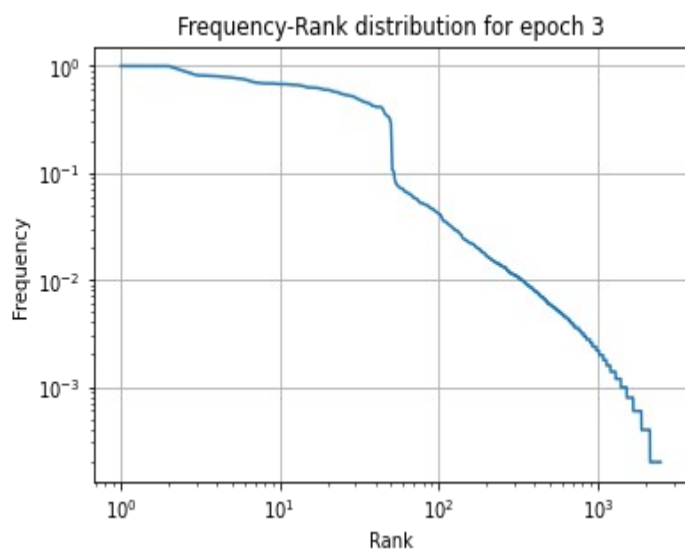
Copy Mutate Algorithm:

- The ratio of Nature Basket size to total number of recipes = 0.16852717855885754
- At epoch 1, KB size = 50, Primordial cuisine size = 500.
- In order to perform mutation the condition of **$\delta \geq \phi$** should be maintained
- **Δ** = $\text{Len}(\text{Kitchen Basket}) / \text{Len}(\text{primordial cuisine})$, **Φ** = $\text{Len}(\text{Nature Basket}) / \text{Len}(\text{Dataset})$

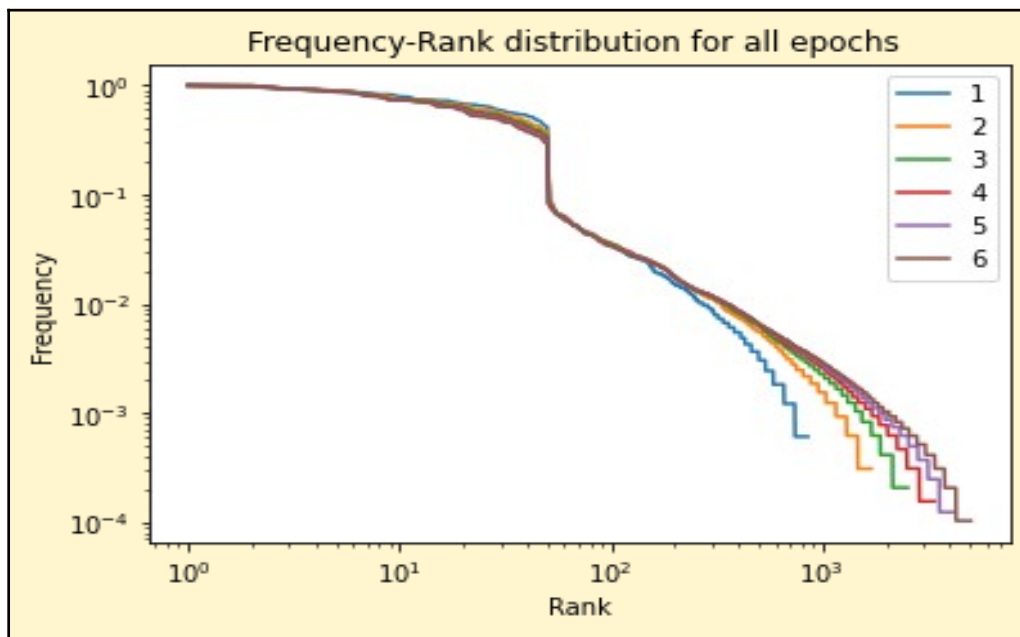
- If this ratio is not maintained then an ingredient is picked from the Nature Basket uniformly randomly and appended to the Kitchen Basket. This is done to maintain the ratio 0.16852717855885754
- Then a recipe is picked from the Primordial cuisine uniformly randomly and an ingredient that is to be replaced is picked from that recipe.
- An ingredient from Kitchen basket is picked and it was checked if the fitness value of the KB ingredient > fitness value of the ingredient to be replaced and the KB ingredient should not be already present in the recipe to be modified.
- Once the recipe is modified then it is appended back to the primordial cuisine. This will continue until we generate 6629 unique recipes in an epoch.
- **Note:** In the first epoch, $6629 - 500 = 6129$ new recipes will be generated followed by 6629 new recipes in the further epochs.

Following are the frequency rank distribution and recipe size distribution graphs of each epoch:





Combined frequency rank distribution graph for the primordial cuisine of recipe size =10



Graph Analysis:

As seen in the graph the frequency rank distribution follows the **power law** but there's some deviation from the power law which is the sharp fall in the graph of each epoch at the rank 10^2 which means that the frequency of usage of ingredients in the cuisine remains somewhat constant initially which may indicate that there are ingredients present in the recipe that have the **fitness value** really high so that they can not be replaced. So, in each cuisine which is made after every epoch doesn't have the uniform distribution of ingredients. There are **some ingredients which highly dominates in the cuisine and rest of the ingredients have comparatively lower usage**.

Also noticed in the python code about this sharp fall in the graph that the trend of usage of ingredients in the cuisine was such that: salt- 60, chilly – 58, onion- 57, oregano – 57, tea- 6, coffee – 3 etc.

The freq rank distribution of initial primordial cuisine follows perfectly the power law because there's no mutation and the ingredients are randomly picked from the KB.

Difference between the cuisine generated after epoch is in the number of ingredients in the epoch. As it can be seen that the graph is widening up as the number of epochs are increasing. This is just because of the number of recipes in the epochs are increasing and thereby increasing the number of different ingredients.

Recipe size distribution graph depicts that there are 100% recipes of size = 10 in each epoch.

Time complexity:

Number of epochs – e

Recipe generated per epochs – n

Total number of recipes generated = $N = n * e$

The upper bound time complexity = $O(N)$

*** Time taken to replace the ingredient until a new recipe is formed is not taken into consideration but in worst case it can be: KB size * recipe size

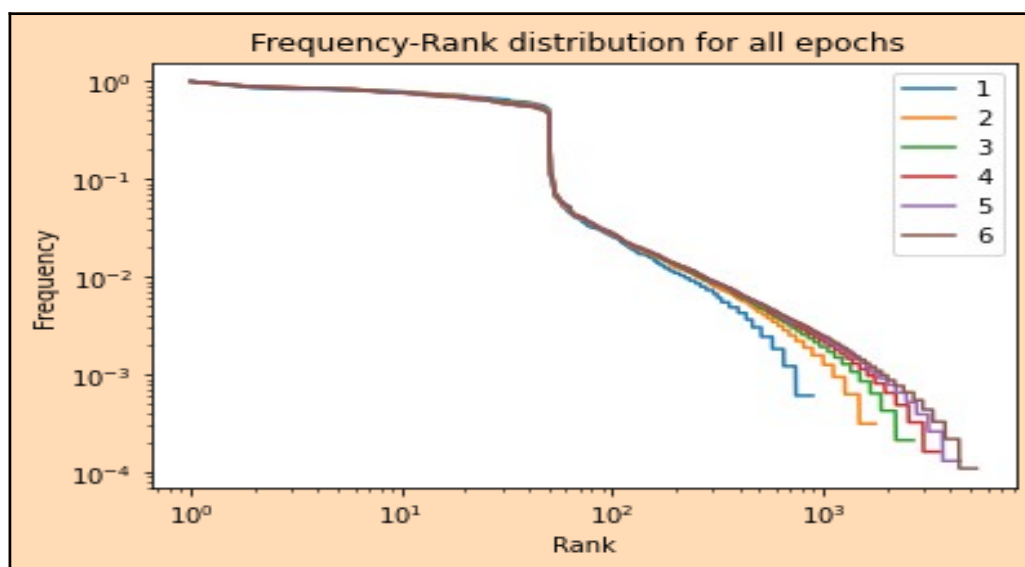
*** Time taken to check if the recipe formed is already present in the Primordial cuisine = **size of primordial cuisine** which is not constant and hence not taken into consideration.

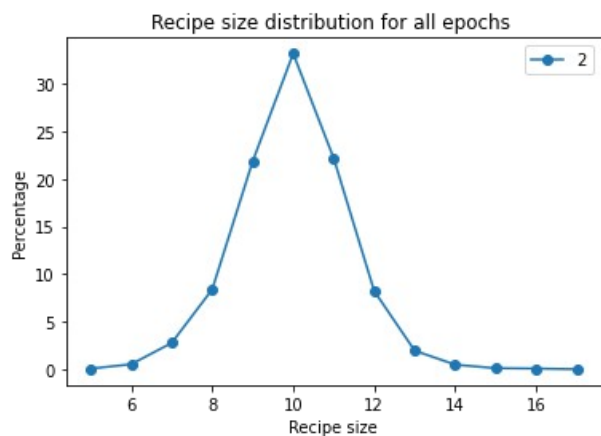
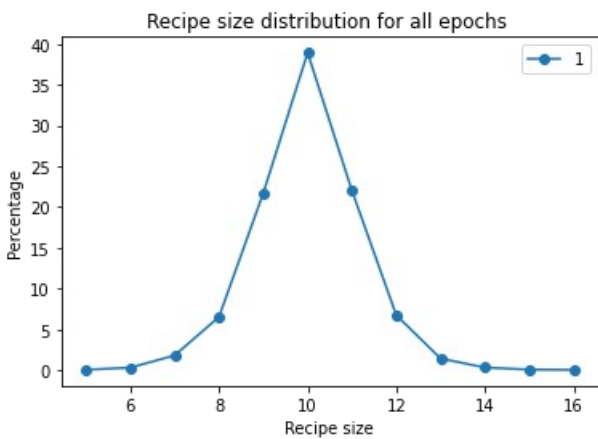
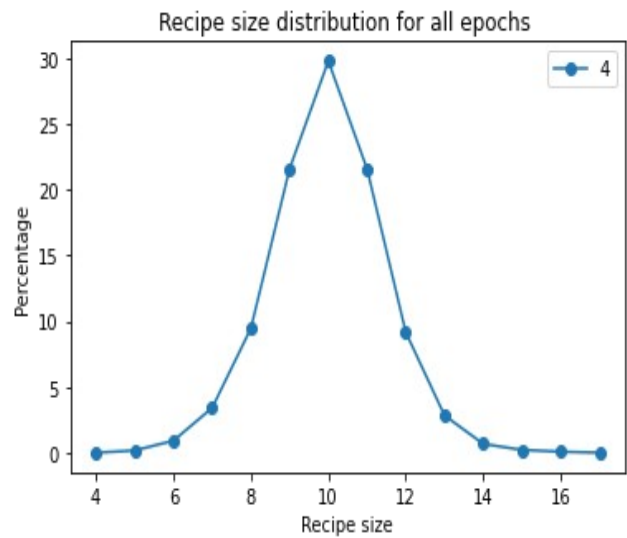
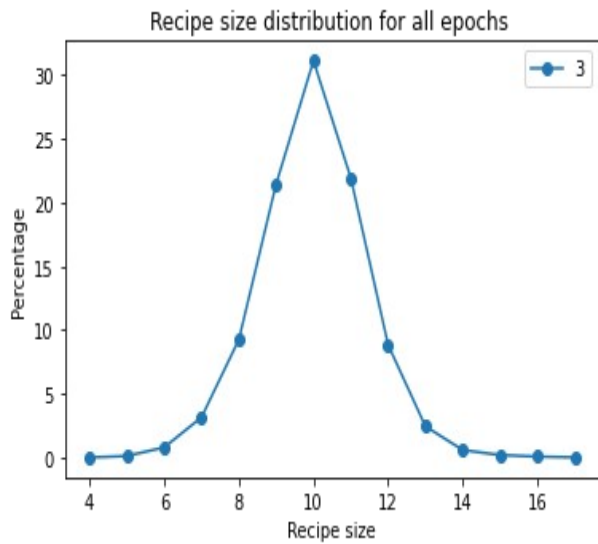
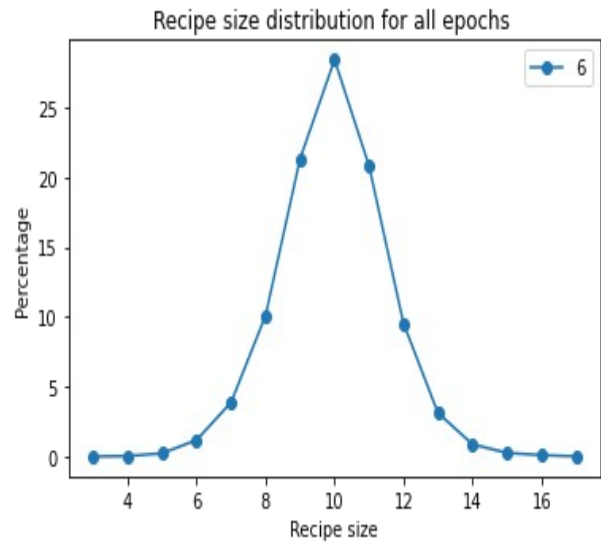
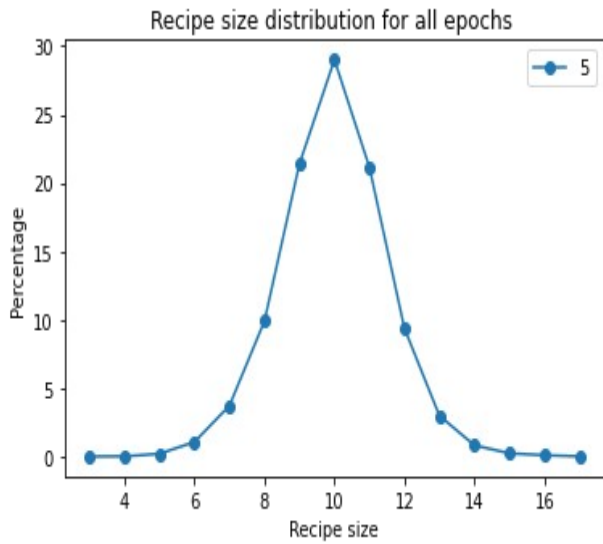
The M ratio is maintained in each epoch which is nearby 0.1685. This is also printed inside the notebook while each epoch is running.

Question 2:

- **Addition** – Probability = 0.25, done randomly
- **Deletion** - Probability = 0.25, done randomly
- **Replacement** – Probability = 0.5, fitness value of ingredient picked from KB is greater than that of the ingredient to be replaced.

Graphs:





Analysis:

Since the probability of replacement is high as compared to deletion and addition therefore the peak of the graph is at Recipe size = 10. There are maximum number of recipes generated for replacement. Overall the graph is normally distributed.

Frequency Rank Distribution:

As seen in the graph the frequency rank distribution follows the **power law** but there's some deviation from the power law which is the sharp fall in the graph of each epoch at the rank 10^2 which means that the frequency of usage of ingredients in the cuisine remains somewhat constant initially which may indicate that there are ingredients present in the recipe that have the **fitness value** really high so that they can not be replaced. So, in each cuisine which is made after every epoch doesn't have the uniform distribution of ingredients. There are **some ingredients which highly dominates in the cuisine and rest of the ingredients have comparatively lower usage.**

Also noticed in the python code about this sharp fall in the graph that the trend of usage of ingredients in the cuisine was such that: salt- 60, chilly – 58, onion- 57, oregano – 57, tea- 6, coffee – 3 etc.

The freq rank distribution of initial primordial cuisine follows perfectly the power law because there's no mutation and the ingredients are randomly picked from the KB.

Difference between the cuisine generated after epoch is in the number of ingredients in the epoch. As it can be seen that the graph is widening up as the number of epochs are increasing. This is just because of the number of recipes in the epochs are increasing and thereby increasing the number of different ingredients.

Time complexity:

Number of epochs – e

Recipe generated per epochs – n

Total number of recipes generated = $N = n * e$

The upper bound time complexity = $O(N)$

*** Time taken to replace the ingredient until a new recipe is formed is not taken into consideration but in worst case it can be: KB size* recipe size

*** Time taken to check if the recipe formed is already present in the Primordial cuisine = **size of primordial cuisine** which is not constant and hence not taken into consideration.

*** Time taken for addition, replacement and deletion is taken constant.