CS301

2022-2023 Spring

Project Report

Group 048

Group Members

Ceren Dinç 28220

Alper Berber 28224

Problem Description

Intuitive Description:

The Bin Packing problem is a well-known optimization problem that requires packing a set of objects of varying sizes into the fewest possible bins (or containers) of a predetermined capacity.
Imagine that you need to fit a specific assortment of items into a specific number of boxes, and that you are aiming to use the least amount of boxes as possible. Each box has a fixed capacity, which means that the volume or weight it can hold is limited. Finding the most effective way to pack the items into the boxes to use the fewest number of boxes is the challenge.
Bin Packing problem can be used in a number of real-world contexts, including logistics, shipping, and storage, where reducing the number of containers used can result in time and resource savings.

Formal Description:

Input:

- A set of n items, where item i has weight w_i, for $i = 1, 2, ..., n$.
- A set of m identical bins, each with capacity c.

Output:

- An integer k that represents the minimum number of bins used.

Constraints:

- For all $i = 1, 2, ..., n$, we have $0 \leq w_i \leq c$.
- For all $j = 1, 2, ..., m$, we have $0 \leq c$.

Objective:

- Minimize k subject to the constraints above.

In other words, we want to find a way to assign each item to a bin such that the total weight of items in each bin does not exceed its capacity c, and we want to use the minimum number of bins possible.

Applications:

Logistics and Transportation:

- The Bin Packing problem is used in the logistics and transportation sectors to optimize the loading of cargo or freight onto vehicles like trucks, ships, and airplanes. The objective is to use the fewest possible containers or vehicles to transport the goods, which can decrease the cost of transportation and boost efficiency.

Manufacturing and Production:

- The Bin Packing problem is used in manufacturing and production processes to optimize the use of raw materials and decrease waste. Manufacturers can decrease the amount of unused materials and save on storage costs by packing the items into minimum number of containers.

Cloud Computing and Resource Allocation:

- The Bin Packing problem is used in cloud computing and resource allocation to optimize the allocation of virtual machines (VMs) among physical servers. Cloud providers can save money on infrastructure costs and power consumption by packing the VMs onto the minimum number of servers.

Cutting and Packing Materials:

- The Bin Packing problem is used to optimize material cutting and packing in sectors like textiles, metalworking, and woodworking to minimize waste and boost productivity.

Stock Management and Inventory Control:

- The Bin Packing problem is employed in the retail and distribution sectors to optimize product storage and distribution in warehouses and retail establishments. Retailers can lower storage costs and increase product availability by packing the products onto the fewest possible shelves or storage spaces.

Hardness of the Problem:

Theorem:

The Bin Packing Problem is NP-hard.

Proof:

The Bin Packing problem is NP-hard, which means that for large problem sizes, computing an optimal solution to this problem is computationally impractical.
In the book "Computers and Intractability: A Guide to the Theory of NP-Completeness" (1979), Garey and Johnson provided one of the earliest proofs of NP-hardness for the Bin Packing problem. By reducing the Subset Sum problem, which is known to be NP-Complete, into the Bin Packing problem, Garey and Johnson demonstrate that the problem is NP-hard. The Subset Sum problem can be reduced to the Bin Packing problem in polynomial time, meaning that the Bin Packing problem is NP-hard.

The Subset Sum Problem:

- Given a set of integers $S = \{s_1, s_2, \ldots, s_{n-1}, s_n\}$ and a target value t,
- Does there exist a subset of S, in which elements sum to t?

Reduction from the Subset Sum Problem:

Assume an instance of the Subset Sum problem is given, $S = \{s_1, s_2, \ldots, s_n\}$ and target value t,
To construct an instance of the Bin Packing problem:

- set the bin capacity to the target value t,
- for each s_i in the set S, create an item with size s_i
- let the number of items be n

It is claimed that reduction from the Subset Sum problem is valid and is an instance of the Bin Packing problem thus constructed is a yes-instance if and only if the Subset Sum problem instance is a yes-instance.

1. Assume that the Subset Sum problem instance has a solution (yes-instance), meaning that a subset S whose elements sum to t exists. Hence, it is possible to use a single bin with capacity t to pack elements in the given subset S.
2. Assume that the Bin Packing problem instance has a solution (yes-instance), meaning that n items can be packed into m bins with capacity t. Hence, it is possible to construct a subset of S by getting integers si which are items packed into the same bin. Since each bin has a capacity of t, integers in the constructed subset are sum to t.

Algorithm Description

Brute Force Algorithm:

```
bins = []
weights = [0.99, 0.01, 0.35, 0.06, 0.27]
best = 0
capacity = 1

def brute_force(index):
    count = 0
    for i in range(len(bins)):
        if bins[i] != 0:
            count += 1
    if index >= len(weights):
        if count < best:
            best = count
        return
    for i in range(len(bins)):
        if bins[i] + weights[index] <= capacity:
            bins[i] += weights[index]
            brute_force(index + 1)
            bins[i] -= weights[index]
```

The algorithm works by iterating over each bin and trying to add the next weight to it. If the weight can fit within the bin's capacity, the weight is added to the bin, and the function is recursively called with the index of the next weight to add. If the weight cannot fit in any bin, a new bin is created, and the function is recursively called with the index of the next weight to add.

The algorithm explores all possible combinations of adding weights to bins and keeps track of the minimum number of bins required to pack all the weights. This brute force algorithm implements the solution for the bin packing problem in $O(n!)$ time.

Heuristic Algorithm:

```
def next_fit(weights):
    best = len(weights)
    bins = []
    current_bin = 0

    for i in range(len(weights)):
        if current_bin == len(bins):
            bins.append(0.0)
        if bins[current_bin] + weights[i] <= 1:
            bins[current_bin] += weights[i]
        else:
            current_bin += 1
            if current_bin == len(bins):
                bins.append(0.0)
            bins[current_bin] += weights[i]

    count = len([i for i in bins if i > 0])
    best = min(best, count)
    return best
```

The heuristic algorithm of the Bin Packing problem works by iterating over each weight in the weights list. If addition of the current weight to the weight of the current bin does not exceed 1, then the current weight is added to the current bin. Else, the current weight is added to the next bin. After the iteration of weights, the number of bins that are filled with at least one item is calculated in order to get the total bin count.

As a greedy algorithm, choices are made locally optimal at each step of the next_fit algorithm. After moving to a new bin, the next_fit algorithm does not reconsider previous bins.

The next_fit algorithm is an approximation algorithm and the approximation ratio is 2.

- Assume weights are sorted in descending order,
- Let next_fit_count be the solution of the next_fit algorithm and optimal_count be the optimal solution.
- In the worst case of the next_fit solution, each bin (except for the last) has a total weight of more than ½. This indicates that the total weight of all items is more than (next_fit_count) / 2.
- Also, optimal solution never exceeds the total weight so total weight is less than or equal to optimal_count.

- The last bin of next_fit_count can have weight less than ½. However, there is always at least one bin which has a weight of more than ½. This shows that optimal_count is at least 2.
- next_fit_count ⩽ 2*optimal_count -1

Algorithm Analysis

Brute Force Algorithm:

Theorem:

Subject to the restriction that each bin has a capacity of 1, the brute_force function for the Bin Packing problem determines the minimum number of bins required to pack a given set of weights.

Proof:

The brute_force function recursively tries all possible combinations of packing items into bins. Each item is either packed into a bin or is under consideration for packing into a bin is maintained as an invariant.

The first step of the algorithm is to determine whether every item has been packed into a bin. If all the items are packed and if fewer bins are used than the current best solution, it updates the best solution so far. If there are items that are not packed, the algorithm tries to pack the next item into each bin in turn, by recursively calling itself with the updated packing.

At each step, the algorithm determines whether the current item can be packed into each bin without exceeding the capacity of 1. If the capacity is not being exceeded, the item is packed into the bin and then the algorithm recursively calls itself with the updated packing. If the capacity is being exceeded, the algorithm tries to pack the item into the following bin.

Proof by Contradiction:

Assumption: The algorithm gives an incorrect solution. There exists a solution which uses fewer number of bins to pack the items.

As explained, the algorithm maintains an invariant that each item is either packed or being considered for packing into a bin. This demonstrates that the algorithm considers all possible packing of the items into bins and will eventually find the minimum number of bins required for the packing. This contradicts the assumption that the algorithm gives an incorrect solution.

By contradiction, it is proven that the brute_force function for the Bin Packing problem determines the minimum number of bins required to pack a given set of weights, subject to the restriction that each bin has a capacity of 1.

The Complexity Analysis:

The worst-case time complexity of the brute_force function is **Θ**(n!), where n is the total number of items.

The brute_force function tries to pack all possible combinations of items in all possible ways.

- The number of possible combinations of n items is n!.
- For each combination, the algorithm tries to pack the items into the available bins, which results in O(n) time.
- Thus, the worst-case time complexity of the brute_force algorithm takes **Θ**(n*n!), and simply **Θ**(n!).

Consider the case where the total number of items is n, each weighing 1, and k bins where k <= n. To pack the first item, any of the k bins can be used. Any remaining k-1 bins can be used for the second item. In each step, the number of available bins decreases by 1 for each item, until the last item is packed into the only available bin. There are *k * (k-1) * (k-2) * ... * 1* possible combinations, which results in k! with only considering k items. Since n > k, all possible combinations of n items should be considered, which is n!. Therefore, the time complexity of the brute_force algorithm is **Θ**(n!).

Heuristic Algorithm:

Correctness Claim:

Given a set of weights, the next_fit function for the Bin Packing problem can pack all items into bins subject to the restriction that each bin has a capacity of 1.

Proof:

The next_fit algorithm iterates over a set of weights and places each item one-by-one. Items are placed into bins considering the restriction that each bin has a capacity of 1. An item is not placed into a bin if the remaining capacity is not enough for its weight.

When there are n items in total, the next_fit algorithm iterates n times by processing each item once.

It can be concluded that the next_fit algorithm can solve the Bin Packing problem.

The Complexity Analysis:

The worst-case time complexity of the next_fit algorithm is O(n), where n is the total number of items.

Θ(n) is a tight upper bound:

- The algorithm iterates over the weights and performs constant time operations for each.
- The algorithm processes each item at least once: Ω(n)
- The algorithm processes each item at most once: O(n)

Sample Generation (Random Instance Generator)

```python
def create_test_case(size: int):
    # construct dataset
    temp_bins = [0.0] * size
    temp_weights = []
    for i in range(size):
        r = float(random.randint(0, 100)) / float(100)
        temp_weights.append(r)
    bins = temp_bins
    weights = temp_weights
    best = int(len(weights))
    return bins, weights, best
```

The code above generates random weights with the given size and creates empty bins. Since the maximum number of bins used can be as many as the size of the *weights* input, the *best* is set to its local maximum *len(weights)*.

Algorithm Implementation

Brute Force Algorithm

The algorithm provided in the *Algorithm Description* part was implemented. Tests were generated with the sample generation function provided in the *Sample Generation* section. Weights generated by the sample generation function are:

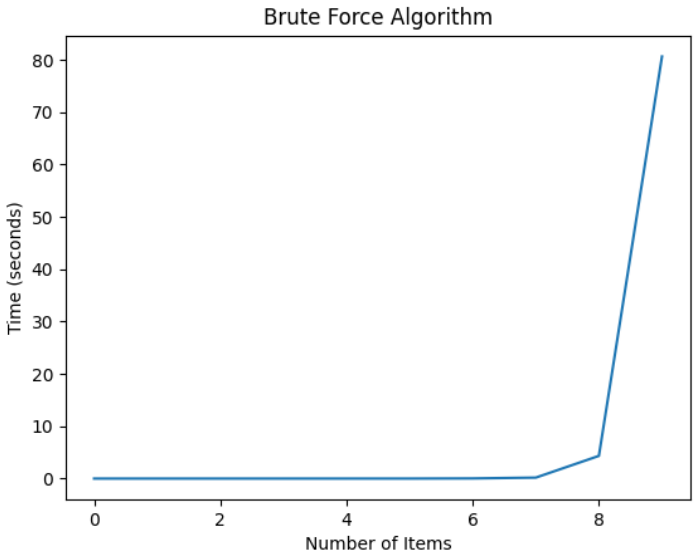
```
weights = [
    [],
    [0.65],
    [0.19, 0.01],
    [0.61, 0.83, 0.56],
    [0.06, 0.39, 0.27, 0.41],
    [0.77, 0.83, 0.68, 0.37, 0.53],
    [0.90, 0.45, 0.07, 0.55, 0.51, 0.44],
    [0.89, 0.65, 0.42, 1.0, 0.07, 0.01, 0.86],
    [0.17, 0.7, 0.43, 0.33, 0.35, 0.77, 0.23, 0.85],
    [0.72, 0.44, 0.1, 0.8, 0.32, 0.17, 0.68, 0.97, 0.24]
]
```

We kept the maximum size for our sample set to 9 because it takes ~45 minutes for 11, and for values bigger than 11 algorithm theoretically takes hours to find the solution.

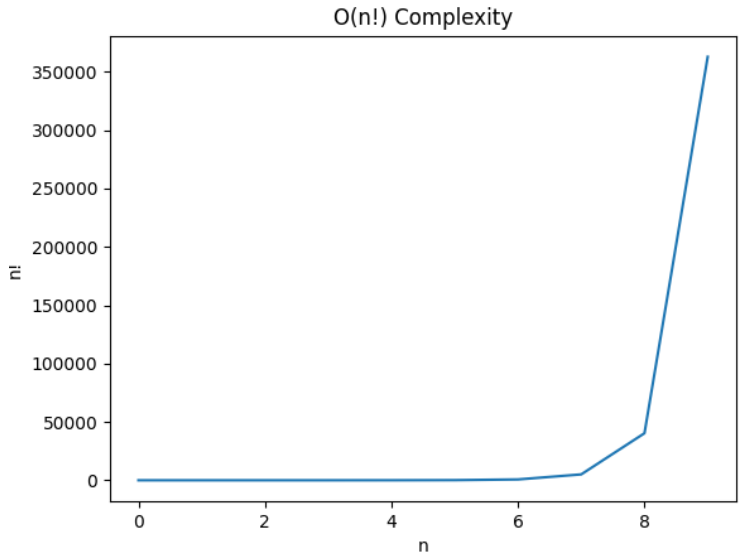
Initial tests of the brute force algorithm results are like this:

```
times = [
    0.000001907349,
    0.000010967255,
    0.000021696091,
    0.000025987625,
    0.000478029251,
    0.000792026520,
    0.027842998505,
    0.182782888412,
    4.333293676376,
    80.639649868011
]
```

Values are in seconds.



This is the graph for the brute force algorithm results in terms of item size and their respective execution times.



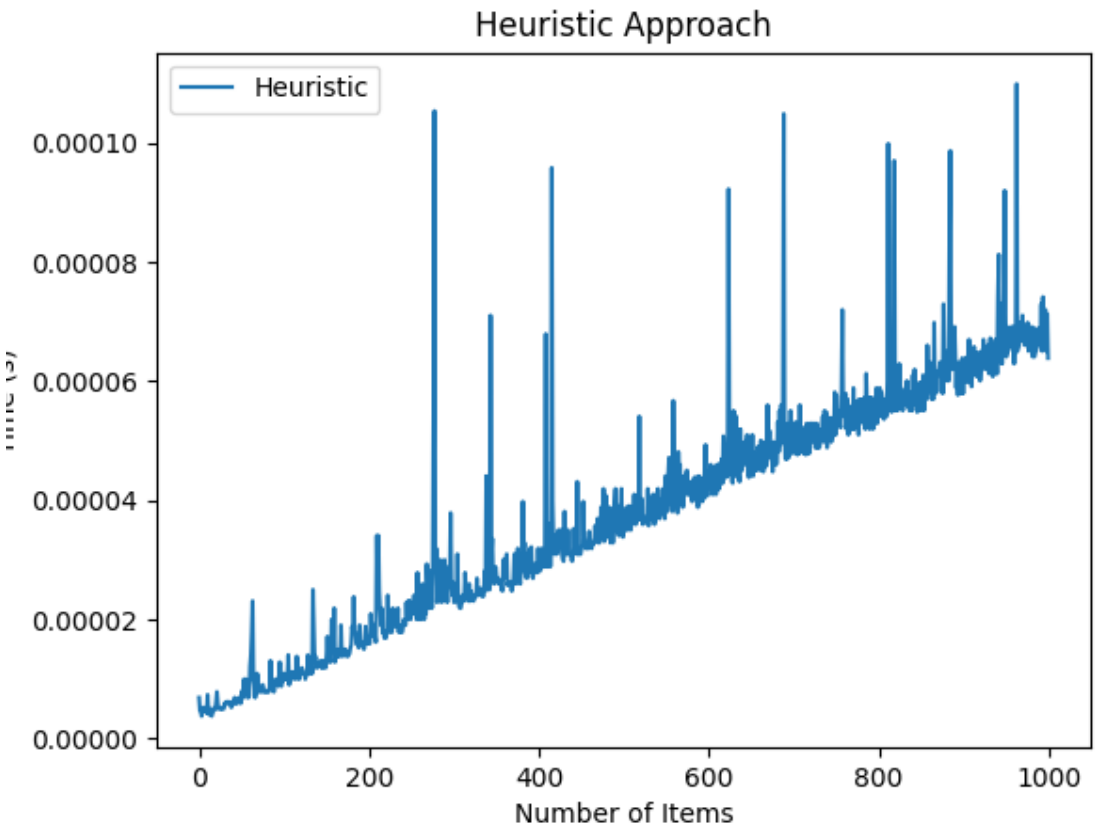
This is the graph of *f(x)=x!* for *0 <= x <= 9*.

As you can see, when two graphs are compared, they look the same. This supports that the brute force algorithm works in O(n!).

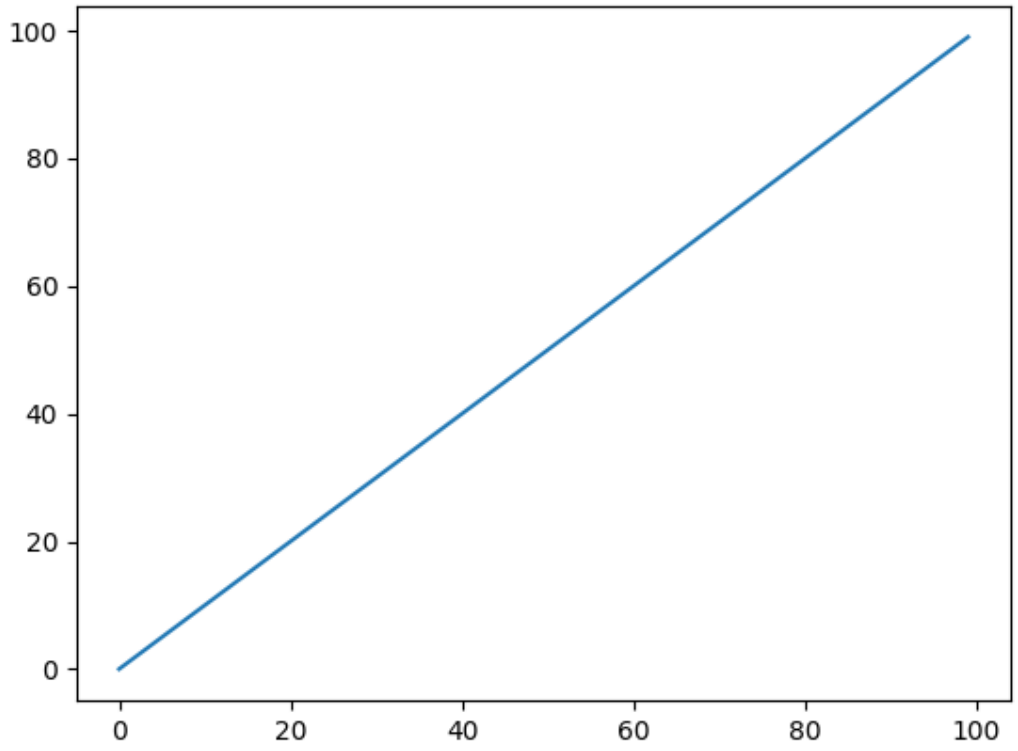
Heuristic Algorithm

The algorithm provided in the *Algorithm Description* part was used as the implementation. Tests were generated with the sample generation function provided in the *Sample Generation* section.

Samples were generated with the function in a range of [0, 100]. The sample size for each size of the weights array is 10.



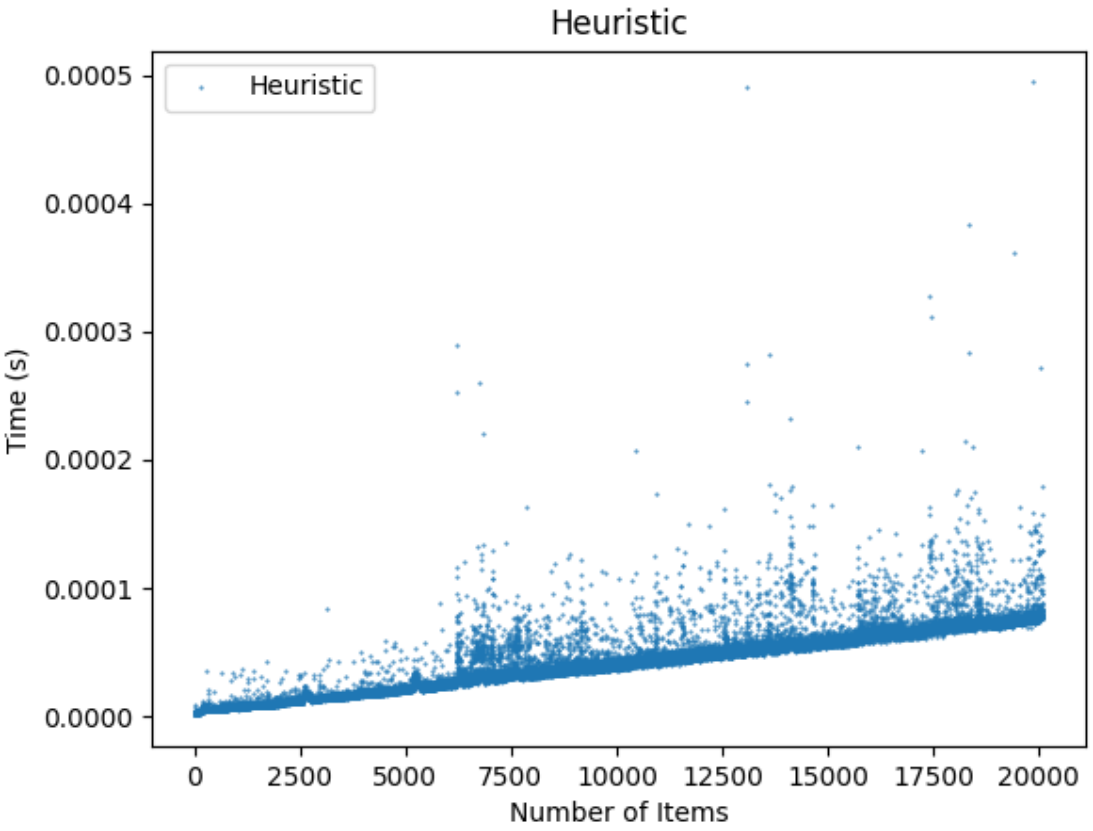
The figure below shows the f(x) = x (linear function) graph. The graph seems similar. This also supports that the heuristic approach for the bin packing problem with the next fit algorithm is *O(n)*.



Experimental Analysis of The Performance (Performance Testing)

The heuristic approach for the bin packing problem can solve the problem in O(n) time. It is remarkably faster when compared to its brute force approach, which is O(n!).

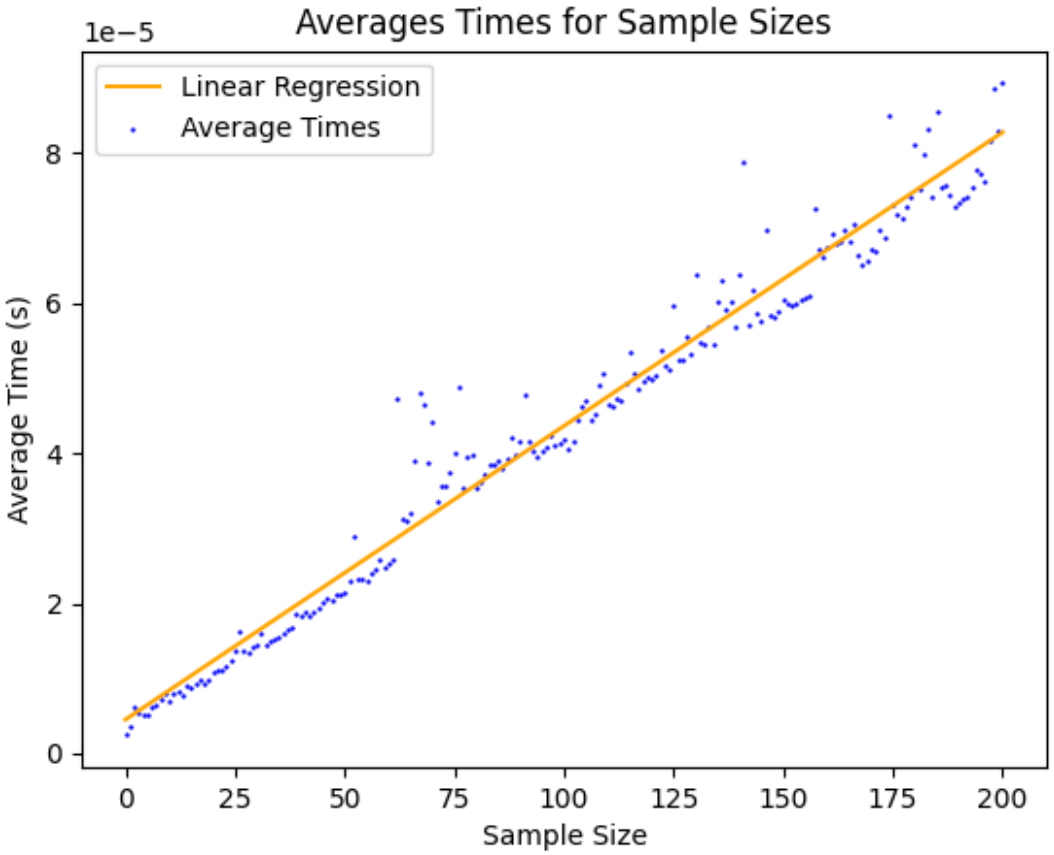
A heuristic approach is used for the following test suites with different test cases in the range [0, 200] elements in the *weights* array. Each size for the *weights* array will be tested with 100 combinations of randomly generated elements.



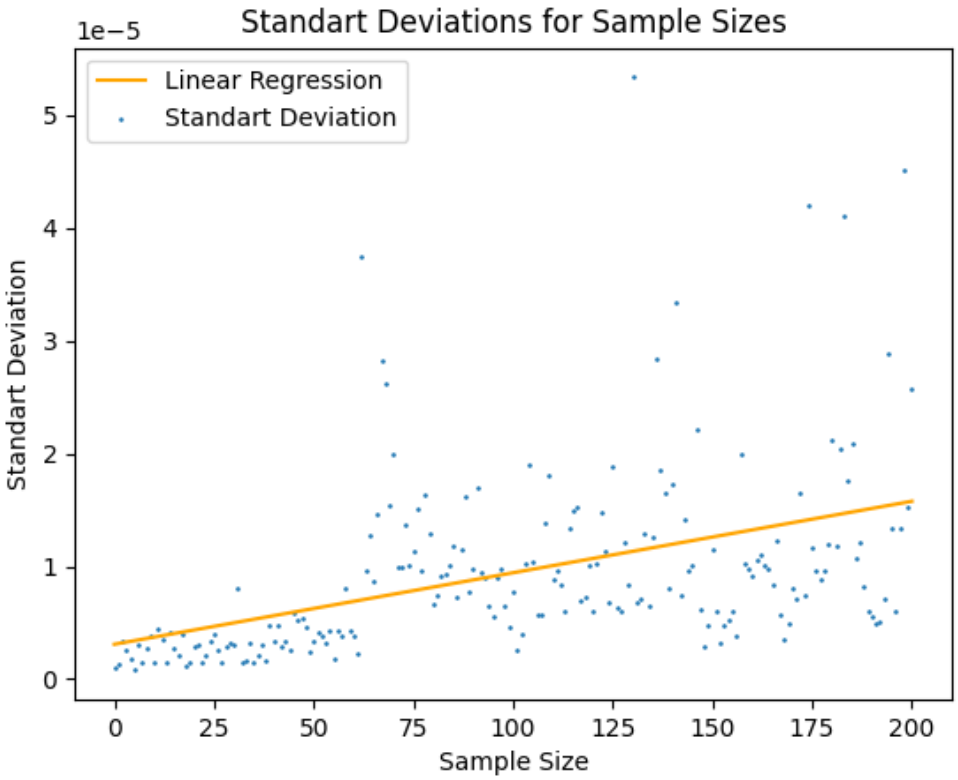
The figure above shows the results of the test run described above. It satisfies the expectation of a function with an O(n) time complexity.

There are a total of **21000** unique runs in this test suite. Since we have run the same sample size with 100 different times, we can get a precise average for each sample size.

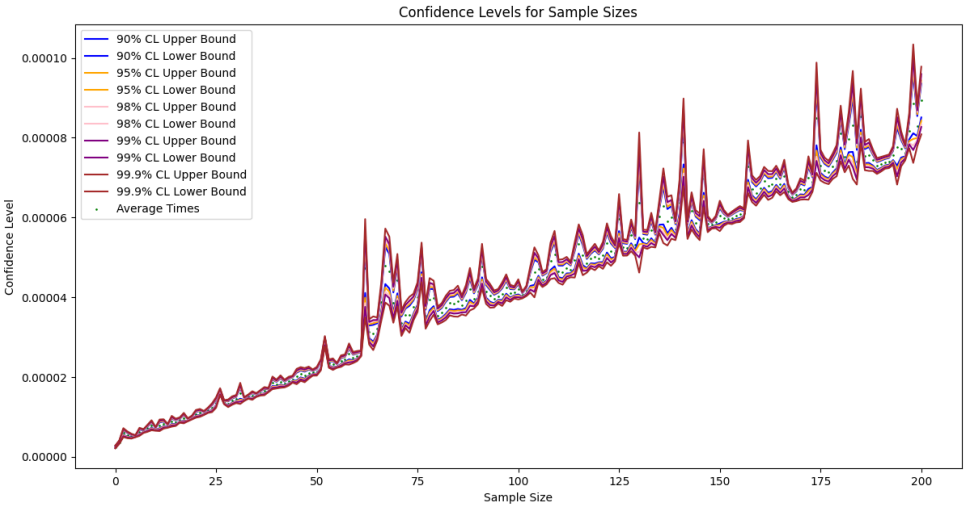
The figure below shows the average times it took for each sample size, with a best-fit line.

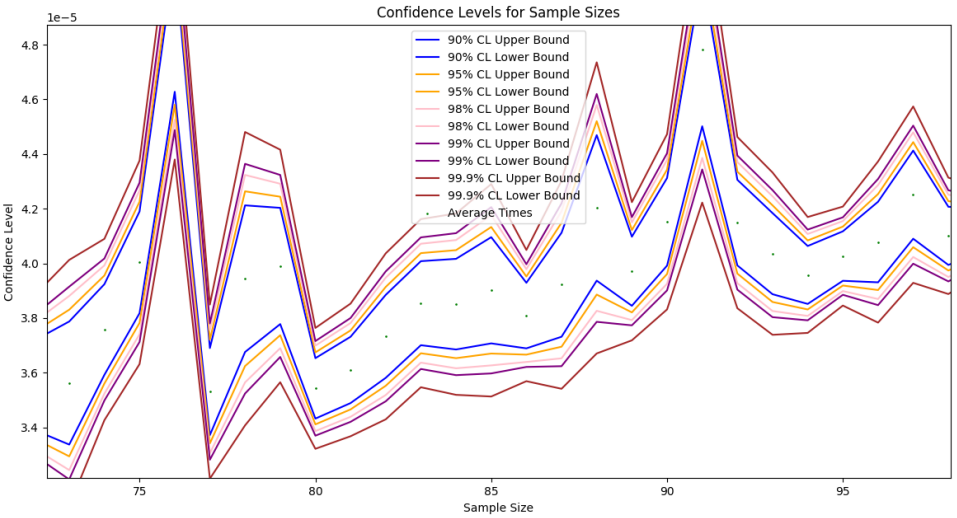


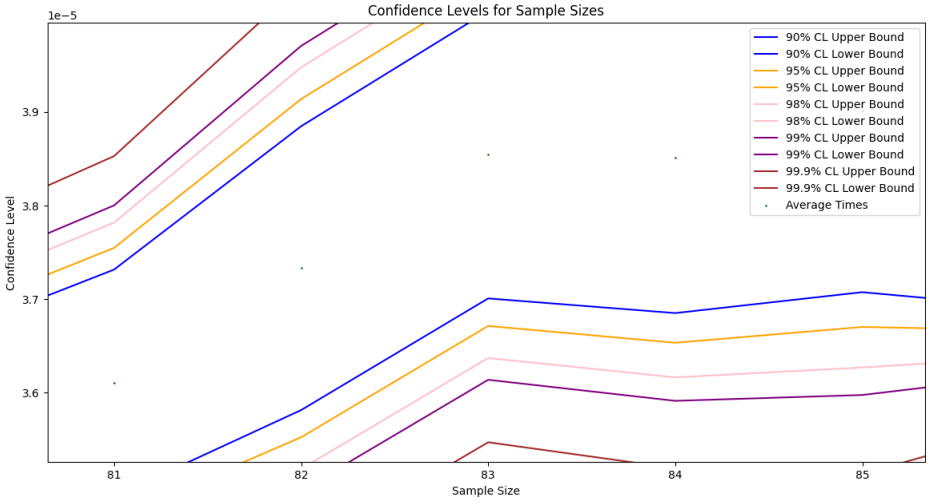
The figure below shows the standard deviation for each sample size:



When we look at the confidence intervals and the average times we get the following figure:

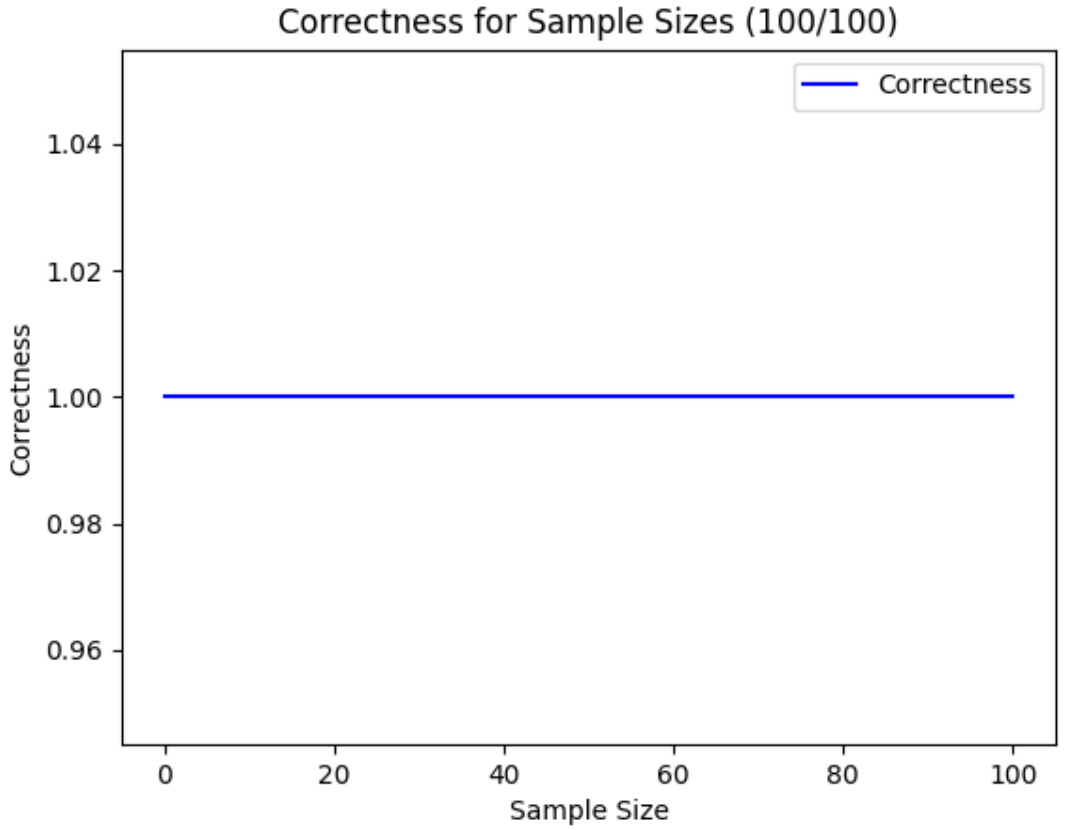






Experimental Analysis of the Quality

In order to assess the quality of the heuristic approach, we've run test suites with [0,9] length *weights* array with 10 repetitions. In total, results of 100 test cases are compared between brute force and heuristic approach.



The heuristic approach gets the same solution as the brute force approach in the sample range.

Experimental Analysis of the Correctness (Functional Testing)

Our comparisons with the brute force approach showed that our heuristic implementation is correct. However, there can be implementation errors. To check for errors in the code, we will use White-Box and Black-Box testing methods to test for edge cases and common algorithm behaviors.

Black Box Testing


```
[
  {
    "weights": [0, 0, 0, 0, 0],
    "results": [
      [5, 0.006140947342],
      [5, 1.3113022e-5]
    ]
  },
  {
    "weights": [1, 1, 1, 1, 1],
    "results": [
      [5, 0.000678777695],
      [5, 1.0967255e-5]
    ]
  },
  {
    "weights": [-1, -1, -1, -1, -1],
    "results": [
      [5, 0.006613969803],
      [5, 1.001358e-5]
    ]
  },
  {
    "weights": [-Infinity, Infinity],
    "results": [
      [2, 9.775162e-6],
      [2, 6.914139e-6]
    ]
  }
]
```


To test some of the edge cases of our algorithm, the starting point will be to try out all 0s or all 1s as our inputs and even negative numbers or infinite values. Even though the behavior for infinite value was undefined in the implementation, it found a solution instead of throwing an error.

White Box Testing

```
def next_fit(self):
   current_bin = 0

   for i in range(len(self.weights)):
      if current_bin == len(self.bins):
         self.bins.append(0.0)
      if self.bins[current_bin] + self.weights[i] <= 1:
         self.bins[current_bin] += self.weights[i]
      else:
         current_bin += 1
         if current_bin == len(self.bins):
            self.bins.append(0.0)
         self.bins[current_bin] += self.weights[i]

   count = len([i for i in self.bins if i > 0])
   best = min(self.best, count)
   return best
```

Since our implementation is not sophisticated, the following case will give us 100% statement coverage.

```
weights = [0.1, 0.4, 0.6, 0.91]
```

This case has the following result:

```
[
  {
    "weights": [0.1, 0.4, 0.6, 0.91],
    "results": [
      [4, 2.145767e-6],
      [4, 1.3113022e-5]
    ]
  }
]
```

Heuristic implementation for the bin packing problem should be implemented correctly, without coding errors.

Discussion

Both the brute force and heuristic algorithms work correctly without any defect. As can be seen from the Functional Testing part, edge cases are covered in the black box testing and %100 statement coverage has been achieved in the white box testing.
There is not any inconsistency between theoretical analysis and experimental analysis. In the algorithm analysis part, we show that the time complexity for brute force algorithm is $\Theta(n!)$ and for the heuristic algorithm is $\Theta(n)$. In the algorithm implementation part, after testing brute force and heuristic algorithms, we also provided graphs for $f(x) = x!$ and $f(x) = x$. Brute force test graph is similar to $f(x) = x!$ and the heuristic test graph is similar to $f(x) = x$ graph.