Project 1: Frequency Bag

CS 0445 (Fall 2014) — Data Structure

Due Friday September 26, 2014 at 11:59pm

The purpose of this project is for you to learn how to implement the Abstract Data Type (ADT) called frequency bag. You will have to understand the behavior of the ADT frequency bag and implement it using **linked list**.

The ADT Frequency Bag

The ADT frequency bag is an unusual ADT to store collection of data. Instead of storing each item individually in this bag, we will store the frequency of each item added into the bag. For this frequency bag, we will only store frequencies of various type of compatible objects. Thus, you must develop this ADT using **generic type**. The following are behavior of our frequency bag:

- Users are only allowed to add entries into the bag but not remove.
- Users can check the number of items inside the bag.
- Users can empty the bag.
- Users can check the number of occurrence (frequency) of a specific item inside the bag.
- Users can check the maximum frequency of items inside the bag.
- Users can check the probability of a specific item inside the bag.
- There is no rule about how items are organized inside the bag.

For example, suppose a user create an empty frequency bag of class wrapper `Integer` named `fb` and adds the following integer into the bag (in that order):

$$3, 6, 7, 1, 5, 2, 3, 7, 4, 7, 9, 5, 3, 7, 6, 4, 7, 4, 2, 5, \text{ and } 3$$

The table below shows the frequency of data added into the bag:

Data	Frequency
1	1
2	2
3	4
4	3
5	3
6	2
7	5
8	0
9	1

The following are returns values of some methods of this frequency bag:

1. `fb.size()` should return 21.
2. `fb.getFrequencyOf(4)` should return 3.
3. `fb.getFrequencyOf(8)` should return 0.
4. `fb.getMaxFrequency()` should return 7 since 7 has the maximum number of occurrences which is 5.
5. `fb.getProbabilityOf(6)` should return 0.095238095 since $2/21 = 0.095238095$.

What to Do?

For this project, you must develop two classes, `FrequencyBag.java` and `CharacterFrequency.java`. Details about each class are as follows:

To Do # 1: `FrequencyBag.java`

For this project, you are going to create a class named `FrequencyBag` which is an implementation of ADT frequency bag. Your class **MUST use a linked list** to store frequencies of data. Start with the starter file named `FrequencyBag` which can be found in our CourseWeb under this project. **Do not change signatures of any methods**.

Note that for this project, you must use linked list. As we discussed in class, linked list has a drawback. Imagine if you develop your class using linked list to store items from the example in previous page, you will end up with a linked list consisting of 21 nodes. Each node for each item. Now, if you need to return the frequency of 3 (`getFrequencyOf(3)`), you need to traverse the linked chain from the first node to the last node and count how many 3 are there. Same situation with the method `getProbabilityOf()`. Imagine if you have to use this frequency bag to store 100,000 or more items. Any applications that use a lot of `getFrequencyOf()` and `getProbabilityOf()` methods will suffer a huge performance lost.

Suppose you store data as a pair of data and frequency (e.g., (1,1), (2,2), (3,4), (4,3), ...). In doing so, the number of nodes in your linked chain to store the same set of data (shown in the example on the previous page) is now only 8. Obviously the methods `getFrequencyOf()` and `getProbabilityOf()` should run a little bit faster. **For this project, you must implement your `frequencyBag` this way because our test class will call methods `getFrequencyOf()` and `getProbabilityOf()` very often**.

Constructor

There will be only one constructor to construct an empty frequency bag as follows:

```
public FrequencyBag()
{
        // TO DO
}
```

Methods

The following are methods from your starter code (`FrequencyBag.java`) that you **must** implement:

```
public class FrequencyBag<T>
{
        public void add(T aData);
        public int getFrequencyOf(T aData);
        public int getMaxFrequency();
        public double getProbabilityOf(T aData);
        public void clear();
        public int size();
}
```

Note that you are **NOT ALLOWED** to use any predefined class in your implementation of `FrequencyBag.java`. However, you are allowed to create any additional private methods or additional classes from scratch (e.g., class `Pair`) for your implementation of `FrequencyBag.java`. Since you must use linked list, for simplicity, create class `Node` as an inner class in your `FrequencyBag.java`.

To Do # 2: `CharacterFrequency.java`

Before you develop this program, it is a good idea to run the provided test class to make sure that your `FrequencyBag.java` works correctly first.

`CharacterFrequency.java` is a program that uses `FrequencyBag.java`. This program will read a provided text file named `letter1.txt` and reports frequency of each character (a to z). The output of this program should look like the following:

```
Character: Frequency
====================
a: 198
b: 17
c: 105
d: 90
e: 330
f: 56
g: 64
h: 134
i: 272
j: 3
k: 9
l: 91
m: 67
n: 245
o: 198
p: 71
q: 3
r: 180
s: 171
t: 251
```

```
u: 68
v: 40
w: 28
x: 6
y: 44
z: 11
```

Note that you **MUST** convert all uppercase characters to lowercase characters before you add them into your frequency bag. For this program, you are allowed to use only predefined classes that are related to reading file and string.

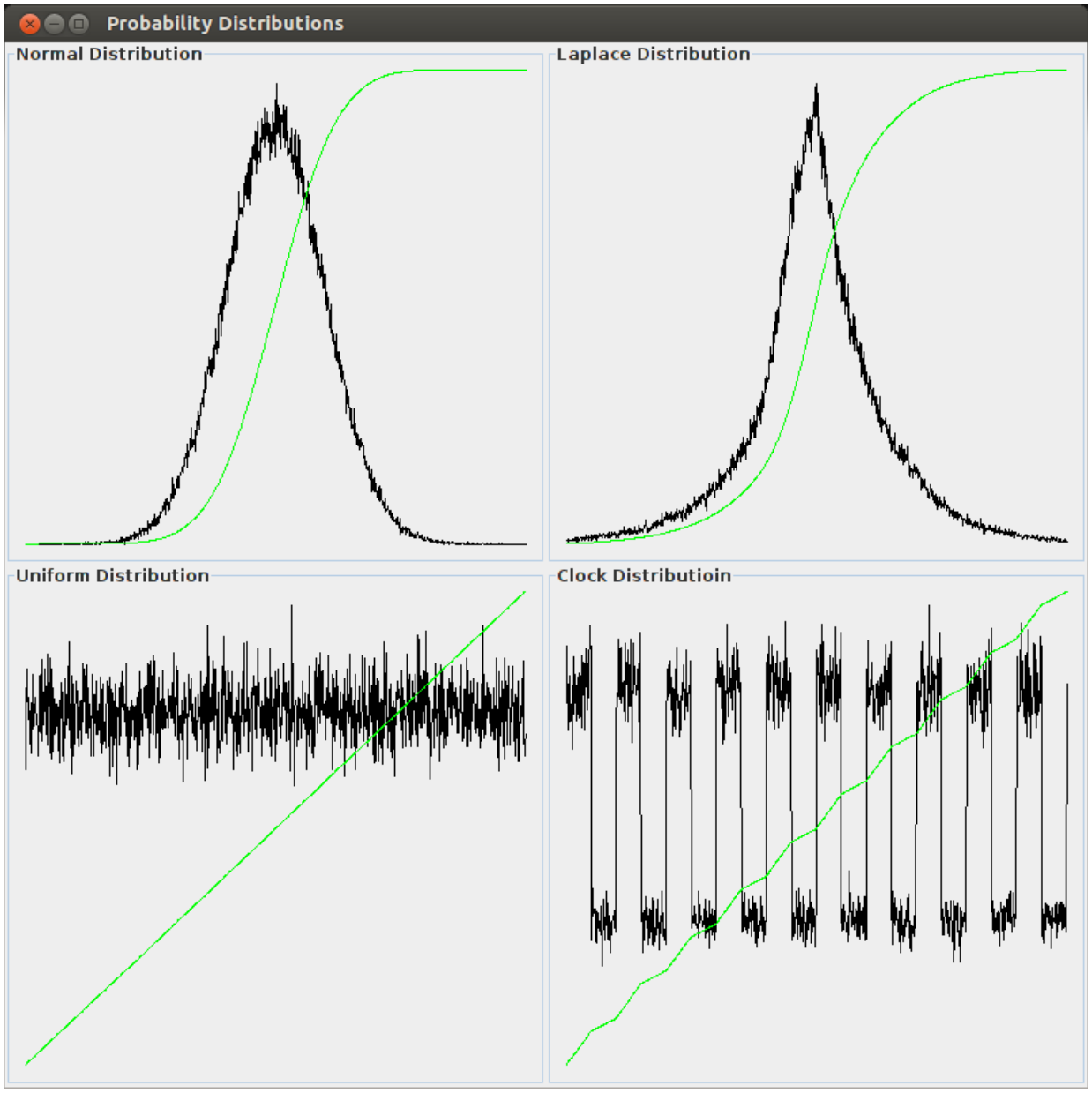
Test Your Class

We supply you two programs to test your class. You must download the following files from CourseWeb into the same directory as your `FrequencyBag.java`:

1. `FrequencyBagTester.java`
2. `FrequencyFrame.java`
3. `FrequencyGraphComponent.java`
4. `RandomDistribution.java`

The test class (`FrequencyBagTester.java`) is provided. Simply compile and run this test class. This program will test your `FrequencyBag` and show results (`PASS`/`FAIL`) for tests. If there are one or more `FAIL`, you should keep trying to fix your code until you get all `PASS`. Note that this test class is not perfect. It cannot tell you why your program is incorrect. You may have to look at the source code of `FrequencyBagTester.java` and see why it says `FAIL` and trace your code.

Another test class is `FrequencyFrame.java`. This class will construct four frequency bags. Each bag will be used to store each sampling type (normal distribution, laplace distribution, uniform distribution, and clock distribution) generated by `RandomDistribution.java`. It also shows progresses of each sampling on a new frame in the form of probability distribution and probability density function. The result should look like the following:



Note that this program will generate around 200,000 data for each sampling type and add them into your frequency bag. Every 10 data added, it will call the method `getFrequencyOf()` 1,000 times and method `getProbabilityOf()` 1,000 times for each sampling type. If you implement your `FrequencyBag` correctly, the performance of this program should be pretty fast. Note that the black lines represents probability distributions and the green line represents probability density functions.

Due Date and Submission

This assignment is due on Friday September 26, 2014 at 11:59pm. **No late submission will be accepted**. Zip all your file (`FrequencyBag.java`, `CharacterFrequency.java` and your additional classes) in to a single file and submit it to the CourseWeb under Project 1.