Predicting the Weather

Lauren Wood
Siena College, Loudonville, NY
(Dated: December 3, 2016)

If there is a 20% chance of it raining on any given day in a month, the probability that it only rains one day is about 0.93%. If there is a 10% chance that it will rain on any given day in a month, the probability that it rains atleast either days is about 0.77%. Given parameters, the probability that it rains at least 10 cm in a given month is about 39%.

I. INTRODUCTION

The point of this project was to use analytic and numeric approaches to calculate difference probabilities of rainfall in a month.

For the numeric portion, the Monte Carlo method was used in python. For the analytic portion, formulas were used and probabilities were calculated by hand. Below, each problem is described in detail.

II. PROBLEM 1

This problem presented the challenge of determining the probability of it raining once and only once in a month, 30 days, when the probability that it rained on any given day was 20%.

The problem was first worked out analytically, finding the probability that it rains only once in a month to be 0.928%. Below are the formula and steps taken to obtain this answer.

$$ndays = 30$$

$$P_{rain} = 0.20$$

$$P_{norain} = 0.80$$

$$P = ((P_{rain})(P_{norain}))(ndays)$$

$$P = ((0.20)(0.80^{29}))(30)$$

$$P = ((0.000309485)(30)) * 100$$

$$P = 0.928\%$$

This problem was then worked out using a Monte Carlo approach in Python to see if the same, or very similar, answer could be obtained. This code is explained below.

To determine the probability it only rained once in a month in python, a function was first created. This function took a random number from 0-1 that was created as input and determined if it was less than 0.20, since this is the percentage of rain on a given day. If it was, the function returned true and if not, false. This is how it was determined if it "rained" on a given day or not.

A nested loop was then used to go over the number of days in a month, 30, and the above function was called to determine the number of days it rained that month. The loop outside of this loop determined that if the number of days it rained was only one out of thirty, then one was added to the number of months it only rained once out of the total. 100,000 months were ran over by the loop.

To determine the probability, the total number of months it only rained once was divided by the number of months being tested and multiplied by 100. This is how the percentage of around 0.953% that it only rains once in a month was determined.

III. PROBLEM 2

This problem asked to find the probability that it rains at least 8 days in a given month when the probability of rain on one day is 10%. This problem was worked out numerically using a Monte Carlo approach and is shown below.

To determine the probability for this scenario, a function was created. This function again took a random number as input and since the probability of it raining in a day was 0.10%, if this random number was less or equal to 0.10, the function returned true. If not, it returned false.

Similarly to problem 1, this function was then called in a nested loop that went over the number of days in a month. If the function was true in this nested loop, then one was added to the total days in a month. The outside function then took the total number of days it rained in a month above and determined that if it was more than or equal to eight days, to add one to the total number of months that it rained eight days. This total number of months that it rained eight days was then divided by the total number of months ran over, which was 100,000 and multiplied by 100. The percentage was determined to be around 0.78%.

IV. PROBLEM 3

A.

This problem stated that if it rains on one day, the odds of a certain amount of rainfall on that same day are:

- 20% for 1 cm
- 30% for 2 cm
- 30% for 3 cm
- 10% for 4 cm
- 10% for 5 cm

Although these are the probabilities of the amount of rain on a particular day, the determination of if it is even going to rain depends on what happened the day, or even many days, before.

If it is the first day of the month, the chance of it raining is 10%. If it rained the day before, but not the day before that, then there is a 20% chance of rain. If it rained both days before, but not the third day before, the chance of rain is 25%. If there was rain for the past three days, or more, there is only a 5% chance of rain. Finally, any other scenario yields a 10% chance of it raining.

The objective was to find the probability that there was at least 10 cm of rain in a given month. This was determined using many different techniques in python and shown below with the Monte Carlo method.

First, a function was created to determine that when it did rain, how much rain fell. This function took in the number of days it rained, and looped over them while determining how much rain fell for each day. The rain for each day was then added to the total amount rained.

To determine the odds of it raining at least 10 cm in a month, a few loops were used. The first loop went over a large number of months while the loop inside of this looped through 30 random numbers, using each to determine when or when it didn't rain based on the percentages above which describe if it has rained any days close to the day being analyzed. Whether it has rained on a given day or not, the loop takes care of this with conditionals. The conditionals make sure to compare the chance of rain on the current day to the chances of rain on past days. This comparing is used to determine the percentage of it raining on the current day. Then, the total number of days it rained was inputted into the function above, which then determined how much rain was produced in each month. If the output, total of this function, was equal to or larger than 10 cm, one was added to the total amount of months.

To find the percentage, the total number of months where it rained at least 10 cm was divided by the total number of months looped over. This number was then multiplied by 100 to convert it into a percentage. Although the code is not functioning properly, the percentage that should have been obtained was between 39% and 40%.

B.

Included here is a histogram of the expected rainfall values for the simulation using the Monte Carlo method determined above.

Since the code for the method above was not functioning properly, the histogram calls the function described earlier to determine the amount of rain produced, in cm, if it rains 15 out of the 30 days in a month. This was then ran over 1000 months. By looking at the graph, it seems that the average amount of rainfall would be about 10-15 cm.

The figure below shows these rainfall values.

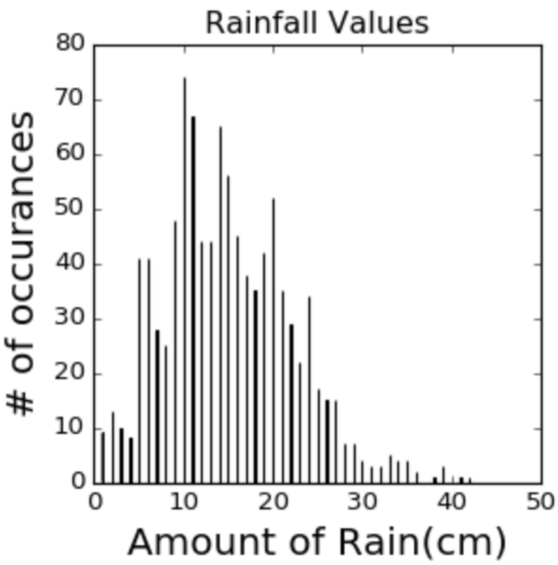

FIG. 1: Expected rainfall values for 1000 months.

C.

The purpose of this question was to obtain the average amount of rainfall in a given month.

The average amount of rain to fall in any given month would be calculated by inputting the number of days it rained out of 30, the number of days in a month, into the function described in part A. Then, taking the output of this function, the total amount of rainfall, and dividing this number by the number of months ran over. The calculation would give the average amount of rainfall in any given month.

Since the code was not functioning properly in part A, and the number of days it rained based on the given parameters was not able to be determined, this average could not be numerically obtained.

D.

The purpose of this section was to find the uncertainty on the average amount of rainfall obtained in part C. Since the average amount of rainfall was not able to be obtained, the uncertainty could not be calculated.

Although the average amount could not be calculated, the way to determine this uncertainty in python is as follows.

First, the middle 95% of the trial values from the Monte Carlo method would be sorted out using numpy/Python sort functions. Then, again using similar sort functions, the high 2.5% of these values would be used to find the high uncertainty and the low 2.5% of these values would be used to find the low uncertainty.

These boundaries then allow for the fact to say that you are 95% confident that the rainfall will be between the high edge value and the low edge value. This makes data much more accurate.

The uncertainty was not able to be determined for this section.