Expected Value and Standard Deviation of the Center of Mass of Random Configurations

A Senior Project submitted to
The Division of Science, Mathematics, and Computing
of
Bard College

by
Nguyen Thuy Linh

Annandale-on-Hudson, New York
May, 2018

Abstract

In his senior project, Finn Hardy determined that the expected value of the center of mass of random configurations on the one-dimensional integer lattice $0, 1, \ldots, n$ is equal to $n/2$, where a random configuration is obtained by randomly assigning to each i between 0 and n a mass of value m or M, with probability p and $1 - p$ respectively. In this project, we will try to find a formula for the standard deviation, and subsequently the distribution, of the center of mass of this lattice. We will use R to create our database and statistically analyze obtained results. So far, we found through our simulations that unlike the expected value, the standard deviation depends on the values of masses, their ratio, the probability p, as well as the number of nodes.

Contents

1
Introduction

1.1 Center of Mass in General

The center of mass is, in simplest words, the mean position of the mass in an object. Due to its widely-used applications, it is not surprising that it piqued attention of many, who set their minds to further analyze the behavior and distribution of the center of mass. There is no doubt in the significance of this concept in sciences such as mathematics or physics. However, one might not realize how omnipresent this idea is in our everyday lives. A few years ago I came across a very interesting type of discipline, namely, rock balancing. It is an art in which rocks or stones of various shapes and sizes are naturally balanced on top of each other without the use of any other supporting materials. Little did I know back then that the stability of the rock structure depends heavily on the location of each stone's center of mass, relative to the support points. Many wonder about how some dancers, for instance, in ballet, seem to defy gravity as they move. The answer lies in the location of one's center of mass, that is, the point where the average distributions of mass of our body is situated in. If you stand straight, assuming a neutral pose, your center of mass will likely be somewhere within your body, most probably below your belly button. However, should you change position of any of your limbs, the center of mass shifts.

For the purpose of this paper, we will consider the center of mass of random configurations on coordinate planes, in which we assign to each position a mass of value m and M, with probability p and $1-p$, respectively. Before we proceed, let us define terms that will be used extensively throughout this project.

Formula for the center of mass:

$$CM = \frac{\sum_{i=0}^{n} iX_i}{\sum_{i=0}^{n} X_i},$$

where n is the number of nodes, i is the index of the location in the lattice, X_i is the random variable for the value of mass, which is m and M with probabilities p and $1-p$ respectively.

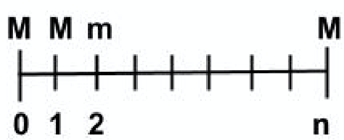
For the sake of our project, we will define a recurring formula as $\alpha_i = m^i + M^i(1-p)$, which, for instance, is present in the expected value of a single X_i: $\mathbb{E}[X_i] = mp + M(1-p) = \alpha_1$.

1.2 Previous Work

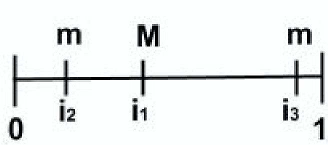
Finn Hardy's theorem says that the expected value for the center of mass on a one-dimensional lattice in the discrete case is $\frac{n}{2}$, where n is the number of nodes.

1.3 Motivation

A few cases we are interested in are the one-dimensional discrete case, one-dimensional continuous case (in which we choose position indices i uniformly), and two-dimensional continuous case (unit circle, in which the angle θ is chosen uniformly). Below are pictures of the first two:



One-dimensional discrete case



One-dimensional continuous case

1.4 Methods

1.5 Summary of Results

2

Center of Mass in the One-dimensional Discrete Case

2.1 Introduction to the Case

We began our project with the investigation into the one-dimensional discrete case. We performed simulations on an integer lattice with positions indexed in order $0, 1, \ldots, n$. As already mentioned, Finn Hardy proved in his senior project that the expected value of the center of mass in this case is $\mathbb{E}[CM] = \frac{1}{2}$. What we are seeking to find is the standard deviation of the center of mass in this system. To prove this, we would first have to find the variance. The difficulty we encountered is that the former term of the typical formula for variance, $\text{Var}(CM) = \mathbb{E}[CM^2] - \mathbb{E}[CM]^2$, i.e., $\text{Var}\left(\frac{\sum_{i=0}^{n} iX_i}{\sum_{i=0}^{n} X_i}\right) = \mathbb{E}\left[\left(\frac{\sum_{i=0}^{n} iX_i}{\sum_{i=0}^{n} X_i}\right)^2\right] - \mathbb{E}\left[\frac{\sum_{i=0}^{n} iX_i}{\sum_{i=0}^{n} X_i}\right]^2$, is very hard to compute. We will discuss the obstacles in detail in sections below.

2.2 Simulations

We created a code in R to conduct simulations for the case considered, using the provided definition of the center of mass. The code is included in Appendix A.1. For the simplicity of experiments, we decided to set the value of $m = 1$, and vary other factors, such as the probability p, number of nodes n, and the value of M. The table below summarizes approximated results for standard deviation of the center of mass obtained through this empirical sampling:

s ->	N = 100		N = 1000		N = 10000	
P	1/M	M/1	1/M	M/1	1/M	M/1
0.1	1	9	3	28	9.5	90
0.2	1.5	6	4.7	19	14.5	58
0.3	2	4	6	14	19	44
0.4	2.5	3.5	7.5	11.5	24	36
0.5	3	3	9	9	29	29
0.6	3.5	2.5	11.5	7.5	36	24
0.7	4	2	14	6	44	19
0.8	6	1.5	19	4.7	58	14.5
0.9	9	1	28	3	90	9.5

Figure 2.2.1. Approximated standard deviation for the one-dimensional discrete case, where s is standard deviation, m = 1 is one of the masses, M is the other mass, p is the probability, and N is the number of nodes.

We notice that, as opposed to the expected value of the center of mass which relies solely on n, the standard deviation depends on all the factors - the probability, the number of nodes, and the ratio of masses. As expected, the results we obtained hint at the proven before fact that the configurations are more or less symmetrical. Intuitively, the values seem to differ by around a factor of 3. Furthermore, if we were to standardize our results (by dividing the position index i by n, thus dividing the results in the table by n), it seems that the standard deviation approaches 0 as n gets larger and larger. We will look into this option of standardization later in the paper. What we will try to do is find a formula for the standard deviation that fits the data obtained through simulations.

2.3 Variance and Standard Deviation

One of the ways we tried to find the standard deviation of the center of mass was by using the common formula for variance, $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. We know from Finn's senior project that the latter term is equal to $\frac{n}{2}$. Now let us look at the former term and apply it to our case. We would have

$$\mathrm{Var}(CM) = \mathbb{E}\left[\left(\frac{\sum_{i=0}^{n} iX_i}{\sum_{i=0}^{n} X_i}\right)^2\right] - \left(\frac{n}{2}\right)^2.$$

The problematic term is the expected value of the center of mass squared. Unfortunately, the numerator and denominator are not independent of each other. Were they to be independent, our calculations would have been greatly simplified. Thus, we had to look for ways of approximating the term, and the formula we came across was second-order Taylor expansion for the expected value of the ratio of two random variables from a paper by CMU. The formula is as follows:

$$\mathbb{E}(R/S) \approx \frac{\mu_R}{\mu_S} - \frac{\mathrm{Cov}(R,S)}{(\mu_S)^2} + \frac{\mathrm{Var}(S)\mu_R}{(\mu_S)^3}.$$

In our case, $R = \left(\sum_{i=0}^n iX_i\right)^2 = \sum_{i=0}^n i^2X_i^2 + \sum_{i\neq j} ijX_iX_j$ and $S = \left(\sum_{i=0}^n X_i\right)^2 = \sum_{i=0}^n X_i^2 + \sum_{i\neq j} X_iX_j$, where $\sum_{i\neq j}$ are double sums $\sum\limits_{i=0}^n \sum\limits_{j=0}^n$ such that $i \neq j$. To calculate single summations, I used Wolfram Alpha online software, and to calculate the more complicated double summation with $i \neq j$, I used Mathematica. I provided the chunk of code in Appendix B. Thus, we have

$$\mathbb{E}[R] = \frac{n(n+1)(2n+1)}{6}\mathbb{E}[X_i^2] + \left[\left[\frac{n(n+1)}{2}\right]^2 - \frac{n(n+1)(2n+1)}{6}\right]\mathbb{E}[X_i]\mathbb{E}[X_j] =$$

$$= \frac{n(n+1)(2n+1)}{6}\alpha_2 + \left[\frac{3n^4+2n^3-3n^2-2n}{12}\right]\alpha_1^2,$$

$$\mathbb{E}[S] = (n+1)\mathbb{E}[X_i^2] + n(n+1)\mathbb{E}[X_i]\mathbb{E}[X_j] = (n+1)\alpha_2 + n(n+1)\alpha_1^2,$$

$$\mathrm{Cov}(R,S) = \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S] = \mathbb{E}\left[\left(\sum_{i=0}^n i^2X_i^2 + \sum_{i\neq j} ijX_iX_j\right)\left(\sum_{i=0}^n X_i^2 + \sum_{i\neq j} X_iX_j\right)\right] - \mathbb{E}[R]\mathbb{E}[S].$$

We analyzed and broke down the former term as follows:

$$\mathbb{E}\left[\left(\sum_{i=0}^n i^2X_i^2 + \sum_{i\neq j} ijX_iX_j\right)\left(\sum_{i=0}^n X_i^2 + \sum_{i\neq j} X_iX_j\right)\right] =$$

$$= \sum_{i=0}^n X_i^4 + \sum_{i\neq j} X_i^3X_j + \sum_{i\neq j} X_i^2X_j^2 + \sum_{i\neq j\neq k} X_i^2X_jX_k + \sum_{i\neq j\neq k\neq l} X_iX_jX_kX_l =$$

$$= (n+1)\alpha_4 + n(n+1)\alpha_3\alpha_1 + 2\cdot\frac{n(n+1)}{2}\alpha_2^2 + 3\cdot\frac{n(n-1)(n+1)}{3}\alpha_2\alpha_1^2 + (2n-n^2+10n^3-15n^4+4n^5)\alpha_1^4.$$

This is incorrect and to be calculated in the future [...]

Another of the ways we tried was to use an approximation directly for the variance by using first-order Taylor expansion from the same paper mentioned above, which is as follows

$$\mathrm{Var}(R/S) = \frac{\mu_R^2}{\mu_S^2}\left[\frac{\sigma_R^2}{\mu_R^2} - 2\frac{\mathrm{Cov}(R,S)}{\mu_R\mu_S} + \frac{\sigma_S^2}{\mu_S^2}\right].$$

In this case, we have $R = \sum_{i=0}^n iX_i$ and $S = \sum_{i=0}^n X_i$. Then:

$$\mathbb{E}[R] = \mathbb{E}\left[\sum_{i=0}^n iX_i\right] = \frac{n(n+1)}{2}\alpha_1,$$

$$\mathbb{E}[S] = \mathbb{E}\left[\sum_{i=0}^{n} X_i\right] = (n+1)\alpha_1,$$

$$\sigma_S = \operatorname{Var}\left(\sum_{i=0}^{n} X_i\right) = \sum_{i=0}^{n} \operatorname{Var}(X_i) = (n+1)(\alpha_2 - \alpha_1^2),$$

$$\sigma_R = \sum_{i=0}^{n} i^2 \operatorname{Var}(X_i) = \frac{n(n+1)(2n+1)}{6}(\alpha_2 - \alpha_1^2).$$

To find the covariance, we used the second-order Taylor expansion to solve for covariance, and more precisely, the term closest to the one we need ($2\frac{\operatorname{Cov}(R,S)}{\mu_S}$):

$$\mathbb{E}(R/S) \approx \frac{\mu_R}{\mu_S} - \frac{\operatorname{Cov}(R,S)}{(\mu_S)^2} + \frac{\operatorname{Var}(S)\mu_R}{(\mu_S)^3}$$

$$\frac{n}{2} = \frac{n}{2} - \frac{\operatorname{Cov}(R,S)}{(\mu_S)^2} + \frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\mu_S^3}$$

$$\frac{\operatorname{Cov}(R,S)}{\mu_S} = \frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\mu_S^2}.$$

Now we will calculate each term of our formula for variance:

$$\frac{\mu_R^2}{\mu_S^2} = \frac{n^2}{4},$$

$$\frac{\sigma_R^2}{\mu_R^2} = \frac{\frac{n^2(n+1)^2(2n+1)^2}{36}(\alpha_2 - \alpha_1^2)^2}{\frac{n^2(n+1)^2}{4}\alpha_1^2} = \frac{(2n+1)^2(\alpha_2 - \alpha_1^2)^2}{9\alpha_1^2},$$

$$2\frac{\operatorname{Cov}(R,S)}{\mu_R\mu_S} = 2\frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\mu_R\mu_S^2} = 2\frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\frac{n(n+1)}{2}\alpha_1(n+1)^2\alpha_1^2} = 2\frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2},$$

$$\frac{\sigma_S^2}{\mu_S^2} = \frac{(n+1)^2(\alpha_2 - \alpha_1^2)^2}{(n+1)^2\alpha_1^2} = \frac{(\alpha_2 - \alpha_1^2)^2}{\alpha_1^2}.$$

Hence, we have

$$\operatorname{Var}(R/S) = \frac{n^2}{4}\left(\frac{(2n+1)^2(\alpha_2 - \alpha_1^2)^2}{9\alpha_1^2} - 2\frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} + \frac{(\alpha_2 - \alpha_1^2)^2}{\alpha_1^2}\right).$$

After simplifying this equation in Mathematica using the simplify function, we get

$$\operatorname{Var}(R/S) = \frac{n^2(\alpha_1^2 - \alpha_2)(9 + \alpha^2(5 + 7n + 4n^2 + 2n^3) - \alpha_2(5 + 7n + 4n^2 + 2n^3))}{18(n+1)\alpha_1^2}.$$

After using FullSimplify, we further obtain

$$\operatorname{Var}(R/S) = \frac{(\alpha_1^2 - \alpha_2)n^2(9 + \alpha_1^2(1+n)(5 + 2n(1+n))) - \alpha_2(1+n)(5 + 2n(1+n))}{18\alpha_1^2(1+n)}.$$

This result yields negative values close to 0, thus it will be reviewed in the future.

2.4 Analysis of Results

We will compare the results obtained from the formula with the ones received from simulations.

2.5 Conclusion

3

Center of Mass in the One-dimensional Continuous Case

3.1 Introduction to the Case and Simulations

We looked at another variation of configurations on the one-dimensional lattice - instead of indexing the positions in order from 0 to n, we chose the indices i uniformly, between 0 and 1. Thus, our formula for the center of mass becomes

$$CM = \frac{\sum_{i=0}^{n} Y_i X_i}{\sum_{i=0}^{n} X_i},$$

where Y_i is a uniform random variable. Unlike the previous case, here the variables X_i and Y_i are independent of each other, and calculating the expected value and standard deviation of the former is not a problem.

We modified the code of the discrete case so that the positions are assigned uniformly. The code is provided in Appendix 1.2. In this case, we will first try to find the expected value and standard deviation of the center of mass "by hand" to see if they match our results from simulations, which will be provided later. In the next chapter we will try to find, or approximate, the expected value of the center of mass in this continuous case, and after that we will look at variance and standard deviation.

3.2 Expected Value

We expect the expected value of the center of mass to be $\frac{1}{2}$, since we the positions are assigned randomly from the uniform function, between 0 and 1 Similarly to the previous case, we still cannot calculate the expected value directly because the numerator and the denominator in $CM = \frac{\sum_{i=0}^{n} Y_i X_i}{\sum_{i=0}^{n} X_i}$ are still dependent on each other. Thus, we decided to use first-order Taylor approximation for this as well. We have

$$\mathbb{E}(R/S) \approx \frac{\mu_R}{\mu_S} - \frac{\text{Cov}(R,S)}{(\mu_S)^2} + \frac{\text{Var}(S)\mu_R}{(\mu_S)^3},$$

where $R = \sum_{i=0}^{n} Y_i X_i$ and $S = \sum_{i=0}^{n} X_i$. Since Y_i is a uniform random variable, we know that $\mathbb{E}[Y_1] = \frac{1}{2}$ and $\text{Var}(Y_i) = \frac{1}{12}$. Since X_i and Y_i are independent, it follows that $\mathbb{E}[X_i Y_i] = \mathbb{E}[X_i]\mathbb{E}[Y_i]$. Thus, we can calculate needed components:

$$\mu_R = \mathbb{E}\left[\sum_{i=0}^{n} Y_i X_i\right] = \sum_{i=0}^{n} \mathbb{E}[Y_i X_i] = \sum_{i=0}^{n} \mathbb{E}[Y_i]\mathbb{E}[X_i] = \frac{1}{2}\mathbb{E}[X_i] = \frac{1}{2}(n+1)\alpha_1,$$

$$\mu_S = \mathbb{E}\left[\sum_{i=0}^{n} X_i\right] = (n+1)\alpha_1,$$

$$\sigma_R = \text{Var}\left(\sum_{i=0}^{n} Y_i X_i\right) = \sum_{i=0}^{n} \text{Var}(Y_i X_i) = \sum_{i=0}^{n} \left[\mathbb{E}[X_i]^2 \text{Var}(Y_i) + \mathbb{E}[Y_i]^2 \text{Var}(X_i) + \text{Var}(Y_i)\text{Var}(X_i)\right] =$$

$$= \sum_{i=0}^{n} \left[\alpha_1^2 \frac{1}{12} + \frac{1}{4}(\alpha_2 - \alpha_1^2) + \frac{1}{12}(\alpha_2 - \alpha_1^2)\right] = \sum_{i=0}^{n} \left[\frac{1}{12}\alpha_1^2 + \frac{1}{4}\alpha_2 - \frac{1}{4}\alpha_1^2 + \frac{1}{12}\alpha_2 - \frac{1}{12}\alpha_1^2\right] =$$

$$= \sum_{i=0}^{n} \left[\frac{1}{3}\alpha_2^1 - \frac{1}{4}\alpha_1^2\right] = (n+1)\left(\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2\right),$$

$$\sigma_S = \text{Var}\left(\sum_{i=0}^{n} X_i\right) = \sum_{i=0}^{n} \text{Var}(X_1) = (n+1)(\alpha_2 - \alpha_1^2),$$

$$\text{Cov}(R,S) = \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S] = \mathbb{E}\left[\sum_{i=0}^{n} Y_i X_i^2 + \sum_{i \neq j}^{n} Y_i X_i X_j\right] - \mu_R \mu_S =$$

$$= \sum_{i=0}^{n} \mathbb{E}[Y_i]\mathbb{E}[X_i^2] + \sum_{i \neq j}^{n} \mathbb{E}[Y_i]\mathbb{E}[X_i X_j] - \mu_R \mu_S = \frac{1}{2}(n+1)\alpha_2 + \frac{1}{2} \cdot 2 \cdot \frac{n(n+1)}{2}\alpha_1^2 - \frac{1}{2}(n+1)^2 \alpha_1^2 =$$

$$= \frac{1}{2}(n+1)\left[\alpha_2 + n\alpha_1 - (n+1)\alpha_1\right] = \frac{1}{2}(n+1)[\alpha_2 - \alpha_1^2].$$

As a result, we obtain the following formula for approximated expected value of the center of mass:

$$\mathbb{E}(R/S) \approx \frac{\frac{1}{2}(n+1)\alpha_1}{(n+1)\alpha_1} - \frac{\frac{1}{2}(n+1)[\alpha_2 - \alpha_1^2]}{(n+1)^2\alpha_1^2} + \frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{1}{2}(n+1)\alpha_1}{(n+1)^3\alpha_1^2} =$$

$$= \frac{1}{2} - \frac{\alpha_2 - \alpha_1^2}{2(n+1)\alpha_1} + \frac{\alpha_2 - \alpha_1^2}{2(n+1)\alpha_1} = \frac{1}{2} - \left(\frac{\alpha_2 - \alpha_1^2 - \alpha_2 + \alpha_1^2}{2(n+1)\alpha_1}\right) = \frac{1}{2} - \left(\frac{\alpha_2 - \alpha_1^2 - \alpha_2 + \alpha_1^2}{2(n+1)\alpha_1}\right) = \frac{1}{2}.$$

Therefore, by the second-order Taylor expansion, the approximated expected value of the center of mass in the continuous case is

$$\mathbb{E}[CM] \approx \frac{1}{2},$$

which is what expect it to be.

3.3 Variance and Standard Deviation

Now we can estimate the standard deviation through calculating variance in two ways; one of them is direct, $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, and the other one is the first-order Taylor expansion, $\mathrm{Var}(R/S) \approx \frac{\mu_R^2}{\mu_S^2}\left[\frac{\sigma_R^2}{\mu_R^2} - 2\frac{\mathrm{Cov}(R,S)}{\mu_R\mu_S} + \frac{\sigma_S^2}{\mu_S^2}\right]$, which is an approximation as opposed to the former. First, we will try using the first formula, which would be in our case:

$$\mathrm{Var}(CM) = \mathbb{E}\left[\left(\frac{\sum_{i=0}^n Y_i X_i}{\sum_{i=0}^n X_i}\right)^2\right] - \mathbb{E}[CM]^2 \approx \mathbb{E}\left[\frac{\left(\sum_{i=0}^n Y_i X_i\right)^2}{\left(\sum_{i=0}^n X_i\right)^2}\right] - \frac{1}{4}.$$

Again, there are two ways of approximating the first term of the above equation. We can either use first-order approximation, which would be $\mathbb{E}[R/S] \approx \frac{\mathbb{E}[R]}{\mathbb{E}[S]}$, or use second-order Taylor expansion. So far, we tried to use the former one. For our case, we have

$$\mathbb{E}[CM] \approx \frac{\mathbb{E}[R]}{\mathbb{E}[S]},$$

where $R = \sum_{i=0}^n Y_i X_i$ and $S = \sum_{i=0}^n X_i$. Thus,

$$\mathbb{E}[CM] \approx \frac{\mathbb{E}[(\sum_{i=0}^n Y_i X_i)^2]}{\mathbb{E}[(\sum_{i=0}^n X_i)^2]},$$

$$\mathbb{E}\left[\left(\sum_{i=0}^n Y_i X_i\right)^2\right] = \sum_{i=0}^n \mathbb{E}[X_i^2]\mathbb{E}[Y_i^2] + \sum_{i\neq j}^n \mathbb{E}[Y_i]\mathbb{E}[Y_j]\mathbb{E}[X_i]\mathbb{E}[X_j],$$

where

$$\mathbb{E}[Y_i^2] = \int_0^1 x^2 \frac{1}{1-0} dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

and

$$\mathbb{E}[X_i^2] = (n+1)\alpha_2 + n(n+1)\alpha_1^2,$$

as calculated in the previous chapter. Thus, we have

$$\frac{\mathbb{E}[(\sum_{i=0}^n Y_i X_i)^2]}{\mathbb{E}[(\sum_{i=0}^n X_i)^2]} = \frac{\frac{1}{3}(n+1)\alpha_2 + \frac{1}{4}\cdot 2 \cdot \frac{n(n+1)}{2}\alpha_1^2}{(n+1)\alpha_2 + n(n+1)\alpha_1^2} = \frac{\frac{1}{3}\alpha_2 + \frac{n}{4}\alpha_1^2}{\alpha_2 + n\alpha_1^2} = \frac{4\alpha_2 + 3n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2} = \frac{1}{4}\left(\frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2}\right).$$

Clearly, $\mathbb{E}\left[(\sum_{i=0}^n Y_i X_i)^2\right] \geq \frac{1}{4}$. If we were to plug this in our approximation for variance, we would get:

$$\mathrm{Var}(CM) \approx \frac{1}{4}\left(\frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2}\right) - \frac{1}{4}.$$

which seems to approach 0 as n gets larger, because the fraction would approach the value of 1.

Also,

$$\mathrm{sd}(CM) = \sqrt{\mathrm{Var}(CM)} = \sqrt{\frac{1}{4}\left(\frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2}\right) - \frac{1}{4}} = \frac{1}{2}\sqrt{\left(\frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2}\right) - 1}.$$

Intuitively, as $n \to \infty$, our result will approach 0, i.e., $n \to \infty$ implies $\mathrm{sd}(CM) \to 0$. Perhaps a thing to consider in the future is to estimate the expected value by using second-order Taylor expansion.

As for the second way to estimate variance, let us look at the first-order Taylor approximation, $\mathrm{Var}(R/S) \approx \frac{\mu_R^2}{\mu_S^2}\left[\frac{\sigma_R^2}{\mu_R^2} - 2\frac{\mathrm{Cov}(R,S)}{\mu_R \mu_S} + \frac{\sigma_S^2}{\mu_S^2}\right]$. Since we calculated needed components in the previous section, we have

$$\mathrm{Var}(R/S) \approx \frac{\frac{1}{4}(n+1)^2\alpha_1^2}{(n+1)^2\alpha_1^2}\left[\frac{(n+1)^2\left(\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2\right)^2}{\frac{1}{4}(n+1)^2\alpha_1^2} - 2\frac{\frac{1}{2}(n+1)[\alpha_2 - \alpha_1^2]}{\frac{1}{2}(n+1)\alpha_1(n+1)\alpha_1} + \frac{(n+1)^2(\alpha_2 - \alpha_1^2)^2}{(n+1)^2\alpha_1^2}\right] =$$

$$= \frac{1}{4}\left[\frac{\left(\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2\right)^2}{\frac{1}{4}\alpha_1^2} - 2\frac{\alpha_2 - \alpha_1^2}{(n+1)\alpha_1^2} + \frac{(\alpha_2 - \alpha_1^2)^2}{\alpha_1^2}\right].$$

The standard deviation would thus be

$$\mathrm{sd}(CM) = \sqrt{\mathrm{Var}(CM)} \approx \sqrt{\frac{1}{4}\left[\frac{\left(\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2\right)^2}{\frac{1}{4}\alpha_1^2} - 2\frac{\alpha_2 - \alpha_1^2}{(n+1)\alpha_1^2} + \frac{(\alpha_2 - \alpha_1^2)^2}{\alpha_1^2}\right]} =$$

$$= \frac{1}{2}\sqrt{\frac{\left(\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2\right)^2}{\frac{1}{4}\alpha_1^2} - 2\frac{\alpha_2 - \alpha_1^2}{(n+1)\alpha_1^2} + \frac{(\alpha_2 - \alpha_1^2)^2}{\alpha_1^2}}.$$

3.4 Simulations

Since we have two approximations for the standard deviation of the center of mass in this continuous case, which are also difficult to simplify and interpret, we conducted simulations to compare the results and check which estimation are suit the data from simulations more (in order to avoid variation of data, I took a mean of five to ten results). On the side, we also tested if the expected value of the center of mass is truly approximately $\frac{1}{2}$ regardless of factors like m, M, n, p, and the simulations confirmed the formula. The code is provided in Appendix 1.3.
We compared different estimation for the standard deviation by calculating absolute errors and relative errors. Below are two of the tables summarizing the data with results for $m = 1$, varied M, $p = 0.5$, and $n = 100$ or $n = 1000$.

P = 0.5, n = 100; m = 1	Standard Deviation			Relative Error		Comparison
M	Simulation (s)	Formula for Variance (s1)	Taylor (s2)	(s - s1) / s1	(s - s2) / s2	(s - s1) / (s - s2)
0.001	0.0300281	0.040383	0.0405817	0.2564173043	0.260058105	0.9811723014
0.01	0.0297754	0.0400321	0.040222	0.25621189	0.2597235344	0.9818218368
0.1	0.0236377	0.0369911	0.0371135	0.3609895353	0.3630969863	0.9909170513
1	0	0.0287242	0.0287242	1	1	1
10	0.0275306	0.0369911	0.0371135	0.2557507076	0.2582052353	0.9872272485
100	0.0294399	0.0400321	0.040222	0.2645926644	0.2680647407	0.9823874755
1000	0.0293061	0.040383	0.0405817	0.2742961147	0.2778493755	0.9823778779

P = 0.5, n = 1000; m = 1	Standard Deviation			Relative Error		Comparison
M	Simulation (s)	Formula for Variance (s1)	Taylor (s2)	(s - s1) / s1	(s - s2) / s2	(s - s1) / (s - s2)
0.001	0.009123482	0.0128842	0.0128906	0.291886031	0.2922375995	0.9983010885
0.01	0.009152822	0.0127702	0.0127764	0.2832671376	0.2836149463	0.998288984
0.1	0.007504018	0.011785	0.011789	0.3632568519	0.3634728985	0.9990665072
1	0	0.00912415	0.00912415	1	1	1
10	0.00745527	0.011785	0.011789	0.3673932966	0.3676079396	0.9990770076
100	0.00903731	0.0127702	0.0127764	0.2923125715	0.2926559907	0.9983418425
1000	0.00920829	0.0128842	0.0128906	0.2853037053	0.2856585419	0.9982619606

It seems that the approximations yield very similar results. The one that uses the typical formula for variance seems to be slightly more accurate in most of the cases. What we noticed is that $n \to \infty \implies \mathrm{sd(CM)} \to 0$.

3.5 Analysis of Results

Although our results suggest that in this continuous case, $\mathbb{E}[CM] \approx \frac{1}{2}$ and $\mathrm{sd}(CM) \to 0$ as n gets larger and larger, we have not yet found a solid proof. It would have been helpful if the approximations had an error term.

We tried to bound the standard deviation instead, to have a better idea of its behavior. It seems that the variance is bounded above by $\frac{1}{12}$, but we do not have enough evidence except from our simulations. One thing we considered is the formula $\mathbb{E}\left[\frac{\left(\sum_{i=0}^{n} Y_i X_i\right)^2}{\left(\sum_{i=0}^{n} X_i\right)^2}\right] - \frac{1}{4}$ - since $0 \leq Y_i \leq 1$, then $\sum_{i=0}^{n} Y_i X_i \leq \sum_{i=0}^{n} X_i$, i.e., $0 \leq \left(\frac{\sum_{i=0}^{n} Y_i X_i}{\sum_{i=0}^{n} X_i}\right)^2 \leq 1$. Thus, we get

$$0 \leq \mathrm{Var}(CM) = \left(\frac{\sum_{i=0}^{n} Y_i X_i}{\sum_{i=0}^{n} X_i}\right)^2 - \frac{1}{4} \leq \frac{3}{4}.$$

However, this result is not yet satisfying. Another option we considered is comparing the standard deviation of the center of mass of the discrete, but standardized case, and the continuous case from this chapter. To standardize the discrete case, we use the formula

$$CM = \frac{\sum_{i=0}^{n} i/nX_i}{\sum_{i=0}^{n} X_i}.$$

Thus, we conducted simulations for the both cases simultaneously, and were able to collect and compare data. We fixed $m = 1$, $M = 10$, $p = 0.5$ and varied number of nodes n. Below is the table with results:

P = 0.5, m = 1, M = 10	Standard Deviation		
N	Discrete Case	Continuous Case	Difference (d-c)
10	0.08203248	0.08125432	0.00077816
50	0.03415174	0.03475454	-0.0006028
100	0.02394078	0.02296108	0.0009797
250	0.015143333	0.01493747	0.000205867
500	0.010383243	0.01064421	-0.000260971
750	0.008711777	0.00859618	0.000115597
1000	0.007517276	0.007378796	0.00013848
5000	0.00333603	0.003300748	0.000035282
10000	0.002326866	0.002312208	0.000014658

I took an average of 10 samplings for better accuracy. As we can see, the standard deviation of the two cases is very similar.

In the future, we will try to find a rigorous proof for expected value and standard deviation in this case by using conditional probability.

3.6 Conclusion

4

Center of Mass in the Two-dimensional Uniform Case

4.1 Introduction to the Case

We were also interested in the two-dimensional uniform case. We looked at a unit circle in an xy-plane with polar coordinates, where the angle θ is determined uniformly. Then we calculated the coordinates (x, y) by setting $x = \cos\theta$ and $y = \sin\theta$. We define the center of mass as follows

$$CM_X = \frac{\sum_{i=0}^{n} \cos(\theta) X_\theta}{\sum_{i=0}^{n} X_\theta},$$

$$CM_Y = \frac{\sum_{i=0}^{n} \sin(\theta) X_\theta}{\sum_{i=0}^{n} X_\theta},$$

We look at the coordinates separately and define $\mathbb{E}[CM] = (x, y)$, $\mathrm{sd}(CM) = (x, y)$. The code used for simulations is included in Appendix A.4. Regardless of the values of m, M, n and p, we got the following results:

$$\mathbb{E}[CM] \approx (0, 0),$$

$$\mathrm{sd}(CM) \approx (0, 0).$$

We will try to prove these in the next sections.

4.2 Expected Value

We will approximate the expected value by first-order approximation, $\mathbb{E}[R/S] \approx \frac{\mathbb{E}[R]}{\mathbb{E}[S]}$. We will first calculate needed components:

$$\mathbb{E}[\cos(\theta)] = \int_0^{2\pi} \cos(t)\frac{1}{2\pi}dt = \frac{\sin(t)}{2\pi}\Big|_0^{2\pi} = 0,$$

$$\mathbb{E}[\sin(\theta)] = \int_0^{2\pi} \sin(t)\frac{1}{2\pi}dt = \frac{-\cos(t)}{2\pi}\Big|_0^{2\pi} = 0,$$

Thus, for the x-coordinate, we have:

$$\mathbb{E}[CM_X] \approx \frac{\mathbb{E}[\sum_{i=0}^{2\pi}\cos(\theta)X_\theta]}{\mathbb{E}[\sum_{i=0}^{2\pi}X_\theta]} = \frac{\sum_{i=0}^{2\pi}\mathbb{E}[\cos(\theta)]\mathbb{E}[X_\theta]}{\sum_{i=0}^{2\pi}\mathbb{E}[X_\theta]} = \frac{\sum_{i=0}^{2\pi}\mathbb{E}[\cos(\theta)]\alpha_1}{n\alpha_1} = 0,$$

$$\mathbb{E}[CM_Y] \approx \frac{\mathbb{E}[\sum_{i=0}^{2\pi}\sin(\theta)X_\theta]}{\mathbb{E}[\sum_{i=0}^{2\pi}X_\theta]} = \frac{\sum_{i=0}^{2\pi}\mathbb{E}[\sin(\theta)]\mathbb{E}[X_\theta]}{\sum_{i=0}^{2\pi}\mathbb{E}[X_\theta]} = \frac{\sum_{i=0}^{2\pi}\mathbb{E}[\sin(\theta)]\alpha_1}{n\alpha_1} = 0.$$

Thus, it follows that $\mathbb{E}[CM] \approx (0, 0)$.

4.3 Variance and Standard Deviation

4.4 Analysis of Results

4.5 Conclusion

5
Theory

I will define: random variable, expected value, variance, standard deviation, covariance. I will include theorems of: variance of two independent random variables.

Appendix A

R Codes

```
# Center of Mass: Discrete Case, one-dimensional

# Simulating
# Use the sample() function to sample from a vector of M and m's.

# Mass values of m and M
m <- 1
M <- 10

# Creating a sample space
sample.space <- c(m,M)

# Probability of getting m
p <- 0.5

# Number of trials
N <- 100

# Create a vector with positions
pos <- 0:(N-1)

# Creating data sets
B <- 10^3-1  # Set number of times to repeat this process
result <- numeric(B)

for(i in 1:B)
{
  index <- sample(sample.space, size = N, replace = TRUE, prob = c(p,1-
p))
  result[i] <- sum(((pos)*index))/sum(index)
}

#result
#m <- mean(result)
s <- sd(result)

paste("sd(CM)=", signif(s))
```

```
# Center of Mass - Continuous Case, one-dimensional

m <- 1 # Mass value of m
M <- 10 # Mass value of M
sample.space <- c(m,M) # Creating a sample space
p <- 0.5 # Probability of getting m
N <- 10^2 # Number of trials

# Continuous case, Simulation

random.numbers <- runif(N, 0, 1) # Vector with positions, uniform from 0
to 1
C <- 10^3-1  # Set number of times to repeat this process
result2 <- numeric(C)

# Creating data sets
for(i in 1:C)
{
  index3 <- sample(sample.space, size = N, replace = TRUE, prob = c(p,1-
p))
  result2[i] <- sum(((random.numbers)*index3))/sum(index3)
}

# Continuous Case, Two different approximations for variance
exsd <- function(M,m,p,n){
  L <- function(pow,i,M,m,p){
    alpha <- (((m^i)*p)+(M^i)*(1-p))^pow
    return(alpha)
  }
  mur <- (1/2)*(n+1)*L(1,1,M,m,p)
  mus <- (n+1)*L(1,1,M,m,p)
  varr <- (n+1)*((1/3)*L(1,2,M,m,p) - (1/4)*L(2,1,M,m,p))
  vars <- (n+1)*(L(1,2,M,m,p)-L(2,1,M,m,p))
  covrs <- (1/2)*(n+1)*(L(1,2,M,m,p)-L(2,1,M,m,p))
  # Calculating the expected value using second-order Taylor expansion:
  ex.taylor <- (mur/mus) - (covrs/mus^2) + (mur*vars/mus^3)
  # Calculating variance using first-order Taylor expansion:
  var.taylor <- (mur^2/mus^2)*((varr/mur^2) - 2*(covrs/(mur*mus)) +
(vars/mus^2))
  sd.taylor <- sqrt(var.taylor)
  # Calculating variance from Var(X)=E[X^2]-E[X]^2 formula:
  var.formula <- (4*L(1,2,M,m,p) + 3*n*L(2,1,M,m,p))/(12*L(1,2,M,m,p) +
12*n*L(2,1,M,m,p)) - (1/4)
  sd.formula <- sqrt(var.formula)
  paste("Taylor: E[CM]=", signif(ex.taylor), "sd(CM)=",
signif(sd.taylor), "Formula: sd(CM)=", signif(sd.formula))
}
paste("Simulation: E[CM]=", signif(mean(result2)), "sd(CM)=",
sd(result2))
exsd(M,m,p,N)
```

```
# Center of Mass - Discrete vs Continuous Case, One-dimensional

# Mass values of m and M
m <- 1
M <- 10

# Creating a sample space
sample.space <- c(m,M)

# Probability of getting m
p <- 0.5

# Number of trials
N <- 10^2

# Vector with positions for the continuous case, uniform from 0 to 1.
random.numbers <- runif(N, 0, 1)

# Vector with positions for the discrete case
pos <- 0:(N-1)

C <- 10^3-1  # Set number of times to repeat this process
resultd <- numeric(C)
resultc <- numeric(C)

# Creating data sets
for(i in 1:C)
{
  index <- sample(sample.space, size = N, replace = TRUE, prob = c(p,1-
p))
  resultd[i] <- sum(((pos/N)*index))/sum(index)
  resultc[i] <- sum(((random.numbers)*index))/sum(index)
}

paste("Discrete: EX(CM)=", signif(mean(resultd)), "Continuous: EX(CM)=",
signif(mean(resultc)))
paste("Discrete: sd(CM)=", signif(sd(resultd)), "Continuous: sd(CM)=",
signif(sd(resultc)))
```

```
# Center of Mass: Two-dimensional, Continuous case, Circle

# Mass values of m and M
m <- 1
M <- 10

# Creating a sample space
sample.space <- c(m,M)

# Probability of getting m
p <- 0.5

# Number of trials
N <- 10^3

# Create a vector with positions
theta <- runif(N, 0, 2*pi)

# Calculating the coordinates
x <- cos(theta)
y <- sin(theta)

# Creating data sets
B <- 10^3-1  # Set number of times to repeat this process
result1 <- numeric(B)
result2 <- numeric(B)

for(i in 1:B)
{
  index <- sample(sample.space, size = N, replace = TRUE, prob = c(p,1-
p))
  result1[i] <- sum((x*index))/sum(index)
  result2[i] <- sum((y*index))/sum(index)
}

#results
m1 <- mean(result1)
m2 <- mean(result2)
s1 <- sd(result1)
s2 <- sd(result2)

paste("E[CM]=(", signif(m1),",", signif(m2), ")")
paste("sd(CM)=(", signif(s1),",", signif(s2),")")
```

Appendix B

Mathematica Codes

```
In[1]:= f[i, j] = i * j

Out[1]= i j

In[2]:= Sum[f[i, j] Boole[i ≠ j], {j, 0, n}, {i, 0, n}]
```

Out[2]= $\frac{1}{12}\left(-2\,n - 3\,n^2 + 2\,n^3 + 3\,n^4\right)$

Bibliography

[1] Howard Zeltman, *Approximations for Mean and Variance of a Ratio*, available at `http://www.stat.cmu.edu/~hseltman/files/ratio.pdf`.