

# Bard

Bard College  
Bard Digital Commons

---

Senior Projects Spring 2018

Bard Undergraduate Senior Projects


---

Spring 2018

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

Thuy Linh Nguyen  
Bard College, [tn3807@bard.edu](mailto:tn3807@bard.edu)

Follow this and additional works at: [https://digitalcommons.bard.edu/senproj\\_s2018](https://digitalcommons.bard.edu/senproj_s2018)

 Part of the [Statistics and Probability Commons](#)



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Nguyen, Thuy Linh, "Expected Value and Standard Deviation of the Center of Mass of Random Configurations" (2018). *Senior Projects Spring 2018*. 292.  
[https://digitalcommons.bard.edu/senproj\\_s2018/292](https://digitalcommons.bard.edu/senproj_s2018/292)

This Open Access work is protected by copyright and/or related rights. It has been provided to you by Bard College's Stevenson Library with permission from the rights-holder(s). You are free to use this work in any way that is permitted by the copyright and related rights. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. For more information, please contact [digitalcommons@bard.edu](mailto:digitalcommons@bard.edu).

Bard

# 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

The goal of this project is to find the expected value and standard deviation of the center of mass in selected random configurations. The center of mass, which is a unique point in a system where the mean distribution of the mass is located, is calculated by dividing the sum of all of the masses times the position they are at by the total mass of the system. The configurations considered in the paper vary upon the way we choose the positions in the configuration. 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  $0, 1, \dots, n$  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, I will propose a formula for the standard deviation of the center of mass of this lattice, as well as the expected value and the standard deviation of the center of mass in two other random configurations: the one-dimensional uniform case, where the positions are chosen uniformly from 0 to 1, and the two-dimensional uniform case, where the angle  $\theta$ , based on whom the  $x$  and  $y$  coordinates are calculated, is chosen uniformly from 0 to  $2\pi$  on a unit circle. RStudio will be extensively used to create our database and statistically analyze obtained results. More complicated computations will be performed in Mathematica.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Dedication</b>	<b>vii</b>
<b>Acknowledgments</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Center of Mass in General . . . . .	1
1.2 Overview . . . . .	2
<b>2 Preliminaries</b>	<b>3</b>
2.1 Motivation . . . . .	3
2.2 Relevant Definitions and Theorems . . . . .	5
2.3 Previous Work . . . . .	9
2.4 Methods . . . . .	11
2.4.1 Formulas . . . . .	11
2.4.2 Algorithms . . . . .	12
2.5 Summary of Results . . . . .	14
<b>3 Center of Mass in the One-dimensional Discrete Case</b>	<b>17</b>
3.1 Introduction to the Case . . . . .	17
3.2 Formula for Variance and Standard Deviation . . . . .	17
3.3 Approximations for Variance and Standard Deviation . . . . .	21
3.4 Simulations and Exact Value . . . . .	25
3.5 Analysis of Results . . . . .	31
<b>4 Center of Mass in the One-dimensional Uniform Case</b>	<b>35</b>
4.1 Introduction to the Case and Simulations . . . . .	35
4.2 Expected Value . . . . .	36
4.3 Formula for Variance and Standard Deviation . . . . .	38

4.4	Approximations for Variance and Standard Deviation . . . . .	44
4.5	Simulations and Exact Value . . . . .	47
4.6	Analysis of Results . . . . .	52
<b>5</b>	<b>Center of Mass in the Two-dimensional Uniform Case</b>	<b>55</b>
5.1	Introduction to the Case . . . . .	55
5.2	Approximation for Expected Value . . . . .	56
5.3	Approximation for Variance and Standard Deviation . . . . .	56
5.4	Simulations . . . . .	57
<b>6</b>	<b>Future Research</b>	<b>59</b>
	<b>Appendices</b>	<b>61</b>
<b>A</b>	<b>R Codes</b>	<b>61</b>
A.1	One-dimensional Discrete Case . . . . .	61
A.2	One-dimensional Uniform Case . . . . .	63
A.3	Standard Deviation: Discrete vs Uniform Case . . . . .	65
A.4	Two-dimensional Uniform Case . . . . .	66
A.5	Sampling Distribution of Sample Means . . . . .	67
<b>B</b>	<b>Mathematica Codes</b>	<b>69</b>
B.1	Algorithm for Calculating Double, Triple and Quadruple Summations . . . . .	69
B.2	Computing the Exact Value Formula for Large $n$ . . . . .	70
	<b>Bibliography</b>	<b>71</b>

# Dedication

To my parents.





# Acknowledgments

First of all, I would like to thank my advisers, Professors Amir Barghi and Maria Belk, for their guidance, commitment and valuable advice throughout this senior project.

I would also like to thank Professors Jim Belk and Stefan Mendez-Diez for their helpful suggestions and ideas to improve the project.

I want to thank my family for their unwavering support and patience, especially my little brother Andy who never fails to make me laugh. A special thanks to my younger sister Weronika who always takes good care of me, and with whom I am happy to have spent two years at Bard.

To all my friends who made my experience at Bard a precious and unforgettable one.

Last but not least, I would like to thank Nyi Nyi, whose amazing sense of humor cheers me up at the most stressful times.



# 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.

Also, locating the center of mass in any system proves to be very useful in many disciplines such as astronomy, body motion, engineering designs, which is why it would be helpful to find any information related to accomplishing this task. In this paper, we will consider simple configurations, in which we will attempt to find the expected value (a predicted value, or a value that the result tends to) and the standard deviation (a measure of how spread out the values are) of the center of mass in hope that any findings will potentially facilitate the analysis of the center of mass in higher-dimensional systems, or more complex configurations.

## 1.2 Overview

This senior project will analyze center of mass in three different configurations separately. First, chapter 2 provides thorough description of the cases considered and the results obtained, definitions and theorems as well as formulas and algorithms used throughout the paper. It will also discuss motivation for choosing this project topic.

Chapters 3, 4 and 5 are each devoted to introducing a new configuration. The chapters have a similar structure: the first sections will briefly describe the cases considered, the next few sections will provide both proofs and approximations for the standard deviation of the center of mass, as well as the expected value of the center of mass (in second and third cases), followed by sections on the simulations performed in the statistical software RStudio in comparison to results obtained from exact value formulas and analysis of results, which includes comparisons of approximations to exact values.

Chapter 6 discusses possible future research that would further explore the topics this senior project addresses.

At the end of the project, appendices with RStudio and Mathematica codes are provided.

# 2

## Preliminaries

The first section of this chapter describes different configurations, whose centers of mass will be analyzed in this project. Next sections provide relevant definitions and theorems used throughout the paper, information about previous work and methods, as well as the summary of results.

### 2.1 Motivation

At the beginning of the project, I was interested in continuing Finn Hardy's senior project [7], which mainly concerned looking for the expectation for the center of mass of finite integer grids, in which the positions are assigned discretely in an orderly manner, so that all points are equally spaced from each other. I tried to find a formula for the standard deviation of the center of mass in such a system. After that I looked at other interesting cases, and attempted to find the expected value and standard deviation in those systems. Thus, in this project, primarily three cases are considered:

- **one-dimensional discrete case**, in which a mass of  $m$  or  $M$  is randomly assigned to points on a one-dimensional lattice that are equally spaced from each other,
- **one-dimensional uniform case**, in which we choose positions uniformly between 0 and 1 and then assign a mass of  $m$  or  $M$  randomly to each position,

- **two-dimensional uniform case**, where we look at a unit circle, the angle  $\theta$  is chosen uniformly, and each point is randomly assigned a mass of  $m$  or  $M$ .

Below are pictures that illustrate the cases of concern. The uniformly chosen positions in the one-dimensional uniform case are labeled as  $y_1, y_2, \dots, y_n$ . The uniformly chosen angles  $\theta_i$  in the two-dimensional case are labeled as  $t_0, t_1, \dots, t_n$ .

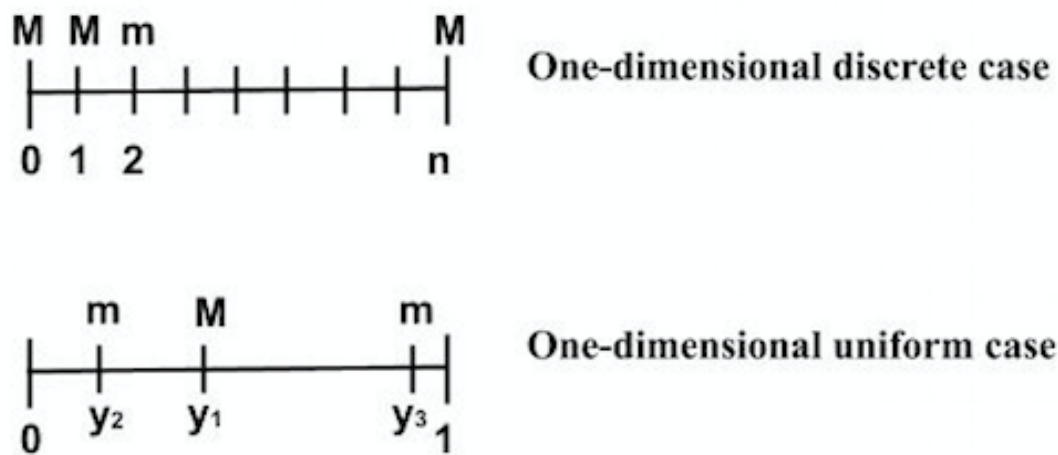


Figure 2.1.1. One-dimensional discrete case (top) and one-dimensional uniform case (bottom)

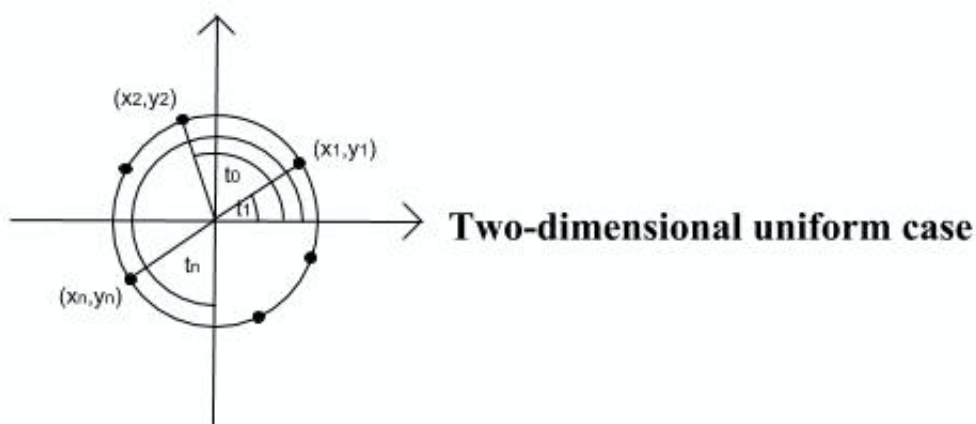


Figure 2.1.2. Two-dimensional uniform case

Due to the widespread applications of the concept of the center of mass in various disciplines, any findings in these cases will hopefully facilitate the process of calculating it in more complicated systems.

## 2.2 Relevant Definitions and Theorems

In this project, 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. An example of such system that we will consider is a one-dimensional lattice with indices from 0 to  $n$  so that all the points are equally spaced from each other. Before we proceed, let us define terms that will be used extensively throughout this project. The theorems and definitions stated can be found in the textbook *Introduction to Probability with Statistical Applications* by Geza Schay [2].

The center of mass is, as mentioned before, is a unique point in a system, where the average distribution of all the masses is located at.

**Definition 2.2.1. (Center of Mass)** The center of mass of a system is defined as

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

where  $n$  is the number of nodes,  $M_i$  is the random variable for the value of mass, which is  $m$  and  $M$  with probabilities  $p$  and  $1 - p$  respectively,  $X_i$  is the coordinate of mass  $M_i$ , and  $i$  is the index numbered between 0 and  $n$ .

In other words,

$$M_i = \begin{cases} m & \text{with probability } p \\ M & \text{with probability } 1 - p. \end{cases}$$

The total mass of the system is given by

$$M_{\text{tot}} = \sum_{i=0}^n M_i.$$

△

First, let us define a random variable:



**Definition 2.2.2. (Random Variable)** A random variable is a real-valued function defined on a sample space.  $\triangle$

For instance, let's say we toss a coin twice, and we observe whether it lands head ( $H$ ) or tail ( $T$ ) up. The sample space in this case is  $\{HH, HT, TH, TT\}$ . The event, say, the “number of heads obtained” is called a random variable, and it is given by a real-valued function

$$X(HH) = 2,$$

$$X(HT) = X(TH) = 1,$$

$$X(TT) = 0.$$

Before defining discrete and continuous random variables, as well as expected value and standard deviation, we need to know what probability and distribution functions are, since their notation will be widely used.

**Definition 2.2.3. (Probability Function)** For any random variable  $X$  on any probability space, the probability function of  $X$  is the function  $f(x) = P(X = x)$ , which is defined for all possible values  $x$  of the random variable  $X$ .  $\triangle$

**Definition 2.2.4. (Distribution Function)** For any random variable  $X$  on any probability space, the distribution function of  $X$  is the function  $F(x) = P(X \leq x)$ , which is defined for all  $x \in \mathbb{R}$ .  $\triangle$

**Definition 2.2.5. (Discrete and Continuous Random Variables)** A random variable is considered to be *discrete* if it has a finite, or a countably infinite number of possible values it can take. A random variable is said to be *continuous* if its possible values constitute a finite or infinite interval. Furthermore, its distribution function is not a step function, but a continuous function.  $\triangle$

**Definition 2.2.6. (Probability Density)** Let  $X$  be a continuous random variable. If there exists a function  $f$  that is nonnegative and integrable over  $\mathbb{R}$ , and for which  $\int_{-\infty}^x f(t)dt = F(x)$  for all  $x$ , then it follows that  $f$  is called the probability function of  $X$ .  $\triangle$

Since we are dealing with expected value and standard deviation of many random variables in this paper, we will define these terms as well as any other related concepts used in the next chapters.

**Definition 2.2.7. (Expected Value)** For any discrete random variable  $X$ , we define the expected value of  $X$  as

$$\mathbb{E}[X] = \sum p_i x_i,$$

where  $p_i = \mathbb{P}(X = x_i)$ , for any finite sums. For any continuous random variable  $X$  with density function  $f(x)$ , we define the expected value of  $X$  as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

as long as the improper integral is absolutely convergent. △

**Theorem 2.2.8. (Expected Value of the Sum of Two Random Variables)** For any two random variables  $X$  and  $Y$  whose expected values exist,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

**Corollary 2.2.9. (Expected Value of a Linear Function of Several Random Variables)**

For any random variables  $X_1, X_2, \dots, X_n$  with finite expected values, where  $n$  is a positive integer, and constants  $a_1, a_2, \dots, a_n$ , it follows that

$$\mathbb{E} \left[ \sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i \mathbb{E}[X_i].$$

**Theorem 2.2.10. (A Constant is Independent of Any Random Variable)** Let  $X$  be any random variable, and let  $Y = a$ , where  $a$  is a constant. Then it follows that  $X$  and  $Y$  are independent of each other.

**Definition 2.2.11. (Independence of Two Random Variables)** Let  $X$  and  $Y$  be random variables.  $X$  and  $Y$  are independent of each other if and only if for all interval  $A$  and  $B$ , the following holds:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

△

**Definition 2.2.12. (Expected Value of the Product of Two Independent Random Variables)** Let  $X$  and  $Y$  be any two independent random variables whose expected value exists. Then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

△

**Definition 2.2.13. (Variance and Standard Deviation)** Let  $X$  be any random variable. Then the variance and standard deviation of  $X$  are defined as follows:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

and

$$\text{SD}(X) = \sqrt{\text{Var}(X)},$$

provided that  $\text{Var}(X)$  exists.

△

**Definition 2.2.14. (Covariance)** Let  $X$  and  $Y$  be any random variables. If  $\mathbb{E}[X]$ ,  $\mathbb{E}[Y]$ , and  $\mathbb{E}[XY]$  all exist, then the covariance of  $X$  and  $Y$  is

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

△

We need to define conditional probability in order to state the theorem of total expectation, which we will use to compute the expected value of the center of mass.

**Definition 2.2.15. (Conditional Expectation for Discrete Random Variables)** If  $X$  and  $Y$  are discrete random variables such that  $f_{X|Y} = \mathbb{P}(X = x | Y = y)$  exists, then the conditional expectation of  $X$  given  $Y = y$  is defined by

$$\mathbb{E}_y[X] = \sum_{x: f_{X|Y}(x, y) > 0} x f_{X|Y}(x, y).$$

△

**Theorem 2.2.16. (*Theorem of Total Expectation*)** *Provided that all expectations below exist, then it follows that*

$$\mathbb{E}[\mathbb{E}_Y[X]] = \mathbb{E}[X].$$

Since  $M_i$  is a random variable that assigns a mass  $m$  with probability  $p$  and a mass  $M$  with probability  $1 - p$ , it follows that

$$\mathbb{E}[M_i] = mp + M(1 - p).$$

As we have to often deal with the expected value of  $M_i$  or  $M_i^2$  in this paper, let us define  $\alpha_a$  as

$$\alpha_a = m^a + M^a(1 - p),$$

where  $a$  is an exponent. We can use this definition as follows:

$$\mathbb{E}[M_i] = mp + M(1 - p) = \alpha_1,$$

$$\text{Var}(M_i) = \mathbb{E}[M_i^2] - \mathbb{E}[M_i]^2 = m^2p + M^2(1 - p) - [mp + M(1 - p)]^2 = \alpha_2 - \alpha_1^2,$$

as well as

$$\mathbb{E}[M_i^a] = m^a p + M^a(1 - p) = \alpha_a.$$

## 2.3 Previous Work

In Finn Hardy's senior project [7], he proved 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. More generally, Hardy proved that for both symmetric and asymmetric configurations, the expected value of the center of mass of any  $n$ -dimensional grid is equal to  $\frac{n}{2}$ . He performed a lot of experiments which theoretically verified his claim that regardless of the values of the masses  $m$  and  $M$ , as well as the probability  $p$ , the center of mass is always in the middle of the grid - the expected value for the center of mass for each component of the lattice turned out to be  $\frac{n}{2}$ . Below are definitions and theorems cited from Hardy's project:

**Definition 2.3.1.** A grid is a coordinate plane consisting of small squares with  $x$ -axis,  $y$ -axis, and  $z$ -axis. Let  $d \geq 1$  and let  $L^d$  be the grid  $[n] \times [n] \times [n] \times \cdots \times [n] = [n]^d$  where  $[n] = \{0, \dots, n\}$ .  $\triangle$

**Theorem 2.3.2.** *The expected value for the center of mass of  $L^1$  is  $\frac{n}{2}$ .*

The following theorem is in Hardy's project, but I will provide a different proof below:

**Theorem 2.3.3.** *Let  $d$  be a positive integer. The expected value for the center of mass for each component of  $L^d$  is  $\frac{n}{2}$ .*

**Proof.** We can prove this claim by using Theorem 2.2.16 about total expectation and conditional probability for discrete random variables. Let  $CM$  be the discrete random variable representing the center of mass, and let the total mass variable, say,  $\text{Mass} = M_{\text{tot}}$ . According to Theorem 2.2.16,  $\mathbb{E}[CM]$  is equal to the expected value of the conditional expected value of  $CM$  given that  $\text{Mass} = M_{\text{tot}}$ . In other words, if we let  $M_{\text{tot}}$  be the total mass of the configuration,  $M_i$  be a random variable that assigns a mass  $m$  with probability  $p$  and a mass  $M$  with probability  $1 - p$ , and  $i$  be the position of the mass  $M_i$  on this one-dimensional lattice, then we would have that

$$\mathbb{E}[CM] = \mathbb{E} \left[ \frac{\sum i M_i}{\sum M_i} \right] = \sum_{k=0}^{n+1} \mathbb{E} \left[ \frac{\sum i M_i}{\sum M_i} \middle| \text{Mass} = M_{\text{tot}} \right] \mathbb{P}(\text{Mass} = M_{\text{tot}}).$$

We will now prove that in the one-dimensional discrete case,

$$\mathbb{E}[CM] = \frac{n}{2}.$$

Let  $k$  be the total number of masses  $m$  in the configuration, and  $n + 1 - k$  be the number of masses  $M$ . Then it follows that

$$M_{\text{tot}} = \sum_{i=0}^n M_i = km + (n + 1 - k)M.$$

We also have that in this case

$$\mathbb{E}[M_i | \text{Mass} = M_{\text{tot}}] = \frac{km + (n + 1 - k)M}{n + 1},$$

$$\mathbb{P}(\text{Mass} = M_{\text{tot}}) = \binom{n + 1}{k} p^k (1 - p)^{n + 1 - k}.$$

Plugging these into our equation, we have

$$\begin{aligned}\mathbb{E}[CM] &= \sum_{k=0}^{n+1} \frac{\mathbb{E}[\sum iM_i | \text{Mass} = M_{\text{tot}}]}{\mathbb{E}[\sum M_i | \text{Mass} = M_{\text{tot}}]} \mathbb{P}(\text{Mass} = M_{\text{tot}}) = \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{km + (n+1-k)M} \cdot \mathbb{E}\left[\sum iM_i | M_{\text{tot}}\right] = \\ &= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{km + (n+1-k)M} \cdot \frac{n(n+1)}{2} \cdot \frac{km + (n+1-k)M}{n+1} = \frac{n}{2}.\end{aligned}$$

□

## 2.4 Methods

### 2.4.1 Formulas

Since the formula for the center of mass is in a fraction form, in calculating its expectation and standard deviation, we need to find ways to compute the expected value of the ratio of two random variables (the numerator and the denominator):

$$\mathbb{E}[CM] = \mathbb{E}\left[\frac{\sum X_i M_i}{\sum M_i}\right].$$

This term proves to be hard to compute unless, for instance, we assume that there are  $k$  of the masses  $m$  and  $n+1-k$  of the masses  $M$ , in which case it would follow that

$$\mathbb{E}[CM] = \mathbb{E}\left[\frac{\sum X_i M_i}{\sum M_i}\right] = \frac{\mathbb{E}[\sum X_i M_i]}{\mathbb{E}[\sum M_i]}.$$

Throughout the paper, Taylor approximation formulas are extensively used to estimate expected value and variance of a ratio of two random variables, which are found in the article by Professor Howard Seltman from CMU [3] who summarizes the resulted approximations using two books: *Kendall's Advanced Theory of Statistics* by Alan Stuart and Keith Ord [4] and *Survival Models and Data Analysis* by Regina C. Elandt-Johnson [5].

Let  $R$  and  $S$  be any random variables. Then the first-order approximation for the expected value of the ratio of two random variables is

$$\mathbb{E}[R/S] \approx \frac{\mathbb{E}[R]}{\mathbb{E}[S]}, \quad (\text{first order})$$

the second-order Taylor approximation for expected value is

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

and the first-order Taylor approximation for variance is

$$\text{Var}(R/S) \approx \frac{\mu_R^2}{\mu_S^2} \left[ \frac{\sigma_R^2}{\mu_R^2} - 2 \frac{\text{Cov}(R, S)}{\mu_R \mu_S} + \frac{\sigma_S^2}{\mu_S^2} \right]. \quad (\text{first order})$$

We can thus estimate the expected value of the center of mass using first two approximations. Once we find the expected value  $\mathbb{E}[CM]$ , we can approximate the standard deviation using the third equation directly, or first using the formula for variance

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

and then either the first-order or second-order Taylor approximation for the expected value of the center of mass squared, after which we can compute standard deviation by taking the square root of the result.

As mentioned above, if we assume the number of one of the masses from the beginning, we can find the expected value and standard deviation without using approximations. Let  $M_i$  be the random variable which is  $m$  with probability  $p$  and  $M$  with probability  $1 - p$ ,  $n$  be the number of nodes,  $k$  be the number of masses  $m$  and  $M_{\text{tot}}$  be the total mass of the system. Then using the theorem for total expectation, we have that the expected value of the center of mass is

$$\mathbb{E}[CM] = \sum_{k=0}^{n+1} \mathbb{E}[CM \mid \text{Total Mass} = M_{\text{tot}}] \mathbb{P}(\text{Total Mass} = M_{\text{tot}})$$

and

$$\mathbb{E}[CM^2] = \sum_{k=0}^{n+1} \mathbb{E}[CM^2 \mid \text{Total Mass} = M_{\text{tot}}] \mathbb{P}(\text{Total Mass} = M_{\text{tot}}),$$

where

$$\mathbb{P}(\text{Total Mass} = M_{\text{tot}}) = \binom{n+1}{k} p^k (1-p)^{n+1-k}$$

and

$$M_{\text{tot}} = \sum_{i=0}^n M_i = km + (n+1-k)M.$$

#### 2.4.2 Algorithms

In order to compute approximations and formulas which include complicated double, triple or quadruple summations, we can use Mathematica. The code for the algorithms is included

in Appendix B. We can also use the *simplify* command in this program to simplify complex equations. Since it is impossible to compute the exact value formulas for big  $n$ 's in R (more specifically, for  $n \leq 10^3 + 28$ ) because of the binomial term inside the summations, Mathematica can be used to do so. Below are calculated summations that are most commonly used throughout the project:

$$\begin{aligned} \sum_{i=0}^n 1 &= n + 1, \\ \sum_{i=0}^n i &= \frac{n(n+1)}{2}, \\ \sum_{i=0}^n i^2 &= \frac{n(n+1)(2n+1)}{6}, \\ \sum_{\substack{i=0 \\ i \neq j}}^n \sum_{j=0}^n 1 &= n(n+1), \\ \sum_{\substack{i=0 \\ i \neq j}}^n \sum_{j=0}^n ij &= \frac{3n^4 + 2n^3 - 3n^2 - 2n}{12} = \frac{n(n+1)(n-1)(3n+2)}{12}. \end{aligned}$$

In this paper,  $\sum_{i=0}^n \sum_{j=0, i \neq j}^n ij$  is abbreviated to  $\sum_{i \neq j}$ . The other summations are calculated in Mathematica analogically to the provided code in Appendix B.

We can perform simulations in RStudio in order to obtain the expected value and standard deviation of the center of mass in different configurations. The codes for the three cases considered in this paper are included in Appendix A. The codes also include algorithms for approximations, which are run simultaneously with the simulations.

In the code for simulations, we set the values of  $m$ ,  $M$ , and create a sample space with these masses using the *sample.space()* command. We then set the value of the probability  $p$  of getting a small mass  $m$ , the number of nodes  $n$ . For the one-dimensional discrete case, we create a simple vector with positions from 0 to  $n$ . For the one-dimensional uniform case, we create a vector with positions chosen randomly and uniformly from 0 to  $n$  by using the *runif()* command. We then assign a mass of  $m$  with probability  $p$  and  $M$  with probability  $1 - p$  using the *sample()* command. As for the two-dimensional uniform case, we create the angles  $\theta$  uniformly from 0 to  $2\pi$ , based on which we obtain  $x$  and  $y$  coordinates of the circle. Then we used the *sample()*



command like in the previous cases, to assign masses to each of the coordinates. We generate data sets and set the number of times to repeat the process of creating a center of mass to be  $10^3 - 1$  times. We then use the *mean()* command to estimate the expected value of the center of mass, and the *sd()* command to find its sample standard deviation. With these data sets, we can also create histograms of results to look at the distribution of the values of the center of mass, or histograms of the sampling distribution of the means.

## 2.5 Summary of Results

Below is the list of theorems and approximations found for expected value and standard deviation of the center of mass in each of the three cases considered. All the formulas have references to the sections which explain how they were derived. Throughout this project, the proven/theoretical formulas are referred to as exact value formulas.

### One-dimensional Discrete Case: Standard Deviation

*Exact Value Formula:*

$$\text{SD}(CM) = \sqrt{-\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}} \quad (\text{section 3.2})$$

*Approximation  $a_1$ :*

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(\alpha_2 + n\alpha_1^2)}} \quad (\text{section 3.3})$$

*Approximation  $a_2$ :*

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(n+1)\alpha_1^2}} \quad (\text{section 3.3})$$

**One-Dimensional Uniform Case: Expected Value***Exact Value:*

$$\mathbb{E}[CM] = \frac{1}{2} \quad (\text{section 4.2})$$

**One-Dimensional Uniform Case: Standard Deviation***Exact Value Formula:*

$$\text{SD}(CM) = \sqrt{\frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}} \quad (\text{section 4.3})$$

*Approximation  $s_1$ :*

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}} \quad (\text{section 4.4})$$

*Approximation  $s_2$ :*

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(n+1)\alpha_1^2}} \quad (\text{section 4.4})$$

**Two-Dimensional Uniform Case:***Expected Value Approximation:*

$$\mathbb{E}[CM] \approx (0, 0) \quad (\text{section 5.2})$$

*Standard Deviation Approximation:*

$$\text{SD}(CM) \approx (0, 0) \quad (\text{section 5.3})$$



# 3

## Center of Mass in the One-dimensional Discrete Case

### 3.1 Introduction to the Case

We begin our project with the investigation into the one-dimensional discrete case. We perform simulations on an integer lattice with positions indexed in order  $0, 1, \dots, n$ . As already mentioned, Finn Hardy proved in his senior project that the expected value of the center of mass in this case is  $\mathbb{E}[CM] = \frac{n}{2}$ . 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 compute the variance. The difficulty here is that the first term of the typical formula for variance,  $\text{Var}(CM) = \mathbb{E}[CM^2] - \mathbb{E}[CM]^2$ , i.e.,  $\text{Var}\left(\frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i}\right) = \mathbb{E}\left[\left(\frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i}\right)^2\right] - \mathbb{E}\left[\frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i}\right]^2$ , is very hard to compute directly. The next section provides a proof for the exact value of the standard deviation of the center of mass, followed by a section discussing possible approximations.

### 3.2 Formula for Variance and Standard Deviation

First, we will try to find the exact value of the variance of the center of mass by using total expectation theorem and conditional expected value. The formula for variance is  $\text{Var}(CM) = \mathbb{E}[CM^2] - \mathbb{E}[CM]^2$ .

**Theorem 3.2.1.** *The variance and the standard deviation of the center of mass in the one-dimensional discrete case are*

$$\text{Var}(CM) = -\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}, \quad (3.2.1)$$

and

$$\text{SD}(CM) = \sqrt{-\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}}. \quad (3.2.2)$$

**Proof.** For the proof, we let  $CM$  be discrete random variable, and let the total mass variable, say,  $\text{Mass} = M_{\text{tot}}$ . According to Theorem 2.2.16,  $\mathbb{E}[CM^2]$  is equal to the expected value of the conditional expected value of  $CM^2$  given that  $\text{Mass} = M_{\text{tot}}$ . In other words, if we let  $M_{\text{tot}}$  be the total mass of the configuration,  $M_i$  be a random variable that assigns a mass  $m$  with probability  $p$  and a mass  $M$  with probability  $1-p$ , and  $i$  be the position of the mass  $M_i$  on this one-dimensional lattice, then we would have that we would need to find the following:

$$\mathbb{E}[CM^2] = \mathbb{E} \left[ \frac{(\sum iM_i)^2}{(\sum M_i)^2} \right] = \sum_{k=0}^{n+1} \mathbb{E} \left[ \frac{(\sum iM_i)^2}{(\sum M_i)^2} \middle| \text{Mass} = M_{\text{tot}} \right] \mathbb{P}(\text{Mass} = M_{\text{tot}}).$$

Let  $k$  be the total number of masses  $m$  in the configuration, and  $n+1-k$  be the number of masses  $M$ . Then it follows that

$$M_{\text{tot}} = \sum_{i=0}^n M_i = km + (n+1-k)M.$$

Given that the total mass of the system is  $M_{\text{tot}}$ , and since we have  $n+1$  nodes, hence

$$\mathbb{E}[M_i^2 \mid \text{Mass} = M_{\text{tot}}] = \frac{km^2 + (n+1-k)M^2}{n+1}.$$

Thus, we have

$$\begin{aligned} \mathbb{E}[CM] &= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot \mathbb{E} \left[ \left( \sum iM_i \right)^2 \middle| M_{\text{tot}} \right] = \\ &= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot \mathbb{E} \left[ \sum_{i=0}^n i^2 M_i^2 + \sum_{i \neq j} ij M_i M_j \middle| M_{\text{tot}} \right]. \end{aligned}$$

Since  $\mathbb{E} \left[ \sum_{i \neq j} ij M_i M_j \mid M_{\text{tot}} \right] = \mathbb{E} [M_i M_j] \sum_{i \neq j} ij$ , we need to find  $\mathbb{E} [M_i M_j \mid \text{Mass} = M_{\text{tot}}]$ , where  $i \neq j$ . Because we are looking at the expected value of the product of two distinct random variables, we can either have the product of two small masses,  $m^2$ , two big masses,  $M^2$ , or one small mass and one big mass ( $mM$  or  $Mm$ ). Given that we have  $k$  of small masses  $m$ , and  $n + 1 - k$  of big masses  $M$ , the number (total count) of such products of two distinct random variables would be  $k(k - 1)$  of  $m^2$ 's,  $(n + 1 - k)(n - k)$  of  $M^2$ 's, and  $2k(n + 1 - k)$  of  $Mm$ 's and  $mM$ 's together. To find the expected value, we have to sum all these possibilities together and then divide the result by  $n(n + 1)$ , which is the total number of combinations we can obtain through taking a product of two distinct random variables out of  $n + 1$  random variables. Thus, we have that

$$\mathbb{E}[M_i M_j \mid \text{Mass} = M_{\text{tot}}] = \frac{mM \cdot 2k(n + 1 - k) + m^2 \cdot k(k - 1) + M^2(n + 1 - k)(n - k)}{n(n + 1)}.$$

With help of Mathematica, I tried to simplify and order the above equation with respect to  $k$ ,  $m$  and  $M$ , and found that it simplifies to the following:

$$\mathbb{E}[M_i M_j \mid \text{Mass} = M_{\text{tot}}] = \frac{(km + (n + 1 - k)M)^2 - (km^2 + (n + 1 - k)M^2)}{n(n + 1)}.$$

As a result,

$$\begin{aligned} \mathbb{E} \left[ \sum_{i \neq j} ij M_i M_j \mid \text{Mass} = M_{\text{tot}} \right] &= \mathbb{E} [M_i M_j] \sum_{i \neq j} ij = \mathbb{E} [M_i M_j] \cdot \frac{3n^4 + 2n^3 - 3n^2 - 2n}{12} = \\ &= \frac{(km + (n + 1 - k)M)^2 - (km^2 + (n + 1 - k)M^2)}{n(n + 1)} \cdot \frac{n(n + 1)(n - 1)(3n + 2)}{12} = \\ &= \frac{(n - 1)(3n + 2)}{12} [(km + (n + 1 - k)M)^2 - (km^2 + (n + 1 - k)M^2)]. \end{aligned}$$

Going back to finding the expected value of the center of mass squared and plugging in the obtained results, we have

$$\begin{aligned} \mathbb{E}[CM^2] &= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n + 1 - k)M)^2} \cdot \mathbb{E} \left[ \sum_{i=0}^n i^2 M_i^2 + \sum_{i \neq j} ij M_i M_j \mid \text{Mass} = M_{\text{tot}} \right] = \\ &= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n + 1 - k)M)^2} \cdot \left( \mathbb{E}[M_i^2 \mid \text{Mass} = M_{\text{tot}}] \sum_{i=0}^n i^2 + \mathbb{E} \left[ \sum_{i \neq j} ij M_i M_j \mid \text{Mass} = M_{\text{tot}} \right] \right) = \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot \left[ \frac{n(n+1)(2n+1)}{6} \cdot \frac{km^2 + (n+1-k)M^2}{n+1} + \right. \\
&\quad \left. + \frac{(n-1)(3n+2)}{12} [(km + (n+1-k)M)^2 - (km^2 + (n+1-k)M^2)] \right] = \\
&= \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot \left[ \frac{(n+1)(n+2)}{12} \cdot (km^2 + (n+1-k)M^2) + \right. \\
&\quad \left. + \frac{(n-1)(3n+2)}{12} \cdot (km + (n+1-k)M)^2 \right] = \\
&= \frac{(n-1)(3n+2)}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot (km^2 + (n+1-k)M^2).
\end{aligned}$$

Hence, the variance of the center of mass would be

$$\begin{aligned}
\text{Var}(CM) &= \mathbb{E}[CM^2] - \mathbb{E}[CM]^2 = \\
&= \left[ \frac{(n-1)(3n+2)}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot (km^2 + (n+1-k)M^2) \right] - \frac{n^2}{4} = \\
&= -\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \cdot (km^2 + (n+1-k)M^2).
\end{aligned}$$

In order to get the standard deviation of the center of mass, we need to take the square root of the equation above. Thus,

$$\text{SD}(CM) = \sqrt{-\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}}.$$

□

Since this formula involves a summation with a binomial inside, it can be hard to compute it for large  $n$ , so it would be better to find a formula that does not include one. Nonetheless, we can use the formula obtained to compute the standard deviation in R for  $n \leq 10^3 + 28$  since R cannot compute the binomial  $\binom{n+1}{k}$  for larger  $n$ 's. It is, however, possible to do so with Mathematica. We tried to simplify the term and it turns out that the *FullSimplify()* function in Mathematica leads to terms containing hypergeometric functions. The output of the command applied to our formula for standard deviation would be

$$\frac{1}{12} \left[ \frac{(n+2)(1-p)^n \Gamma\left(\frac{m+Mn}{m-M}\right)^2 \left((n+1)^2 p(m+M) {}_3\tilde{F}_2\left(-n, \frac{m+Mn}{m-M}, \frac{m+Mn}{m-M}; \frac{2m+M(n-1)}{m-M}, \frac{2m+M(n-1)}{m-M}; \frac{p}{p-1}\right)\right)}{m-M} \right]$$

$$\left. - \frac{(p-1)(m-M) {}_3\tilde{F}_2 \left( -n-1, \frac{M(n+1)}{m-M}, \frac{M(n+1)}{m-M}, \frac{m+Mn}{m-M}, \frac{m+Mn}{m-M}; \frac{p}{p-1} \right)}{m-M} - n - 2 \right],$$

where  ${}_3\tilde{F}_2$  is a type of the generalized, or Gauss's hypergeometric function [8] defined by the hypergeometric series as follows

$${}_p\tilde{F}_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \cdot \frac{z^n}{n!},$$

where  $(x)_n = x(x-1) \cdots (x-(n-1))$  for  $n \geq 0$  is a falling factorial. As we can see, hypergeometric functions contain summations as well, which might not solve our original issue.

### 3.3 Approximations for Variance and Standard Deviation

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

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

The problematic term is the expected value of the center of mass squared. Unfortunately, the numerator and denominator are not independent of each other. Thus, we need to look for ways of finding or approximating the term. We could either use first-order approximation for the expected value of the ratio of two random variables:

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

or second-order Taylor expansion for the expected value of the ratio of two random variables from a paper by Professor Howard Seltman from CMU [3]. The second-order Taylor approximation the expected value of a ratio of two random variables is as follows:

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

We can use any of these to estimate  $\mathbb{E}[CM^2]$ , since  $\mathbb{E}[CM]$  was already computed.



First, we will use the first-order approximation. We will prove that the standard deviation of the center of mass is:

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(\alpha_2 + n\alpha_1^2)}}. \quad (3.3.1)$$

To show that this is true, we want to calculate the expected value of the center of mass squared using the approximation:

$$\mathbb{E}[CM^2] = \mathbb{E} \left[ \frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i} \right] \approx \frac{\mathbb{E} [\sum_{i=0}^n iM_i]}{\mathbb{E} [\sum_{i=0}^n M_i]}.$$

We calculate the nominator and denominator separately:

$$\begin{aligned} \mathbb{E} \left[ \sum_{i=0}^n iM_i \right] &= \mathbb{E} \left[ \sum_{i=0}^n i^2 M_i^2 + \sum_{i \neq j} ij M_i M_j \right] = \mathbb{E} [M_i^2] \sum_{i=0}^n i^2 + \mathbb{E}[M_i M_j] \sum_{i \neq j} ij = \\ &= \mathbb{E} [M_i^2] \sum_{i=0}^n i^2 + \mathbb{E}[M_i] \mathbb{E}[M_j] \sum_{i \neq j} ij = \frac{n(n+1)(2n+1)}{6} \alpha_2 + \frac{n(n+1)(n-1)(3n+2)}{12} \alpha_1^2. \\ \mathbb{E} \left[ \sum_{i=0}^n M_i \right] &= \mathbb{E} \left[ \sum_{i=0}^n M_i^2 + \sum_{i \neq j} M_i M_j \right] = \mathbb{E} [M_i^2] \sum_{i=0}^n 1 + \mathbb{E}[M_i M_j] \sum_{i \neq j} 1 = \\ &= \mathbb{E} [M_i^2] \sum_{i=0}^n 1 + \mathbb{E}[M_i] \mathbb{E}[M_j] \sum_{i \neq j} 1 = (n+1)\alpha_2 + n(n+1)\alpha_1^2. \end{aligned}$$

Thus, we have

$$\begin{aligned} \mathbb{E}[CM^2] &\approx \frac{\mathbb{E} [\sum_{i=0}^n iM_i]}{\mathbb{E} [\sum_{i=0}^n M_i]} = \frac{\frac{n(n+1)(2n+1)}{6} \alpha_2 + \frac{n(n+1)(n-1)(3n+2)}{12} \alpha_1^2}{(n+1)\alpha_2 + n(n+1)\alpha_1^2} = \\ &= \frac{1}{12} \frac{2n(n+1)(2n+1)\alpha_2 + n(n+1)(n-1)(3n+2)\alpha_1^2}{(n+1)\alpha_2 + n(n+1)\alpha_1^2} = \frac{2n(2n+1)\alpha_2 + n(n-1)(3n+2)\alpha_1^2}{12(\alpha_2 + n\alpha_1^2)}. \end{aligned}$$

Hence, we have the

$$\mathbb{E}[CM] \approx \frac{1}{12} \frac{2n(2n+1)\alpha_2 + n(n-1)(3n+2)\alpha_1^2}{\alpha_2 + n\alpha_1^2}.$$

Then the variance of the center of mass would be

$$\begin{aligned} \text{Var}(CM) &= \mathbb{E}[CM^2] - \mathbb{E}[CM]^2 \approx \frac{1}{12} \frac{2n(2n+1)\alpha_2 + n(n-1)(3n+2)\alpha_1^2}{\alpha_2 + n\alpha_1^2} - \frac{n^2}{4} = \\ &= \frac{1}{12} \left[ \frac{2n(2n+1)\alpha_2 + n(n-1)(3n+2)\alpha_1^2 - 3n^2(\alpha_2 + n\alpha_1^2)}{\alpha_2 + n\alpha_1^2} \right] = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{12} \left[ \frac{(4n^2 + 2n - 3n^2)\alpha_2 + (3n^3 - n^2 - 2n - 3n^3)\alpha_1^2}{\alpha_2 + n\alpha_1^2} \right] = \\
&= \frac{1}{12} \left[ \frac{(n^2 + 2n)\alpha_2 - (n^2 + 2n)\alpha_1^2}{\alpha_2 + n\alpha_1^2} \right] = \frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(\alpha_2 + n\alpha_1^2)}.
\end{aligned}$$

The approximation of the standard deviation, which we will call  $a_1$ , is therefore

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(\alpha_2 + n\alpha_1^2)}}. \quad (a_1)$$

We can now try to compute the variance using mentioned second-order Taylor approximation for expected value of the ratio of two random variables. In our case,  $R = (\sum_{i=0}^n iM_i)^2 = \sum_{i=0}^n i^2 M_i^2 + \sum_{i \neq j} ij M_i M_j$  and  $S = (\sum_{i=0}^n M_i)^2 = \sum_{i=0}^n M_i^2 + \sum_{i \neq j} M_i M_j$ , where  $\sum_{i \neq j}$  are double sums  $\sum_{i=0}^n \sum_{j=0}^n$  such that  $i \neq j$ . We can calculate all the the summations in Mathematica.

The code is in Appendix B.1. Thus, we have

$$\begin{aligned}
\mathbb{E}[R] &= \frac{n(n+1)(2n+1)}{6} \mathbb{E}[M_i^2] + \left[ \left[ \frac{n(n+1)}{2} \right]^2 - \frac{n(n+1)(2n+1)}{6} \right] \mathbb{E}[M_i] \mathbb{E}[M_j] = \\
&= \frac{n(n+1)(2n+1)}{6} \alpha_2 + \left[ \frac{3n^4 + 2n^3 - 3n^2 - 2n}{12} \right] \alpha_1^2,
\end{aligned}$$

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

$$\text{Cov}(R, S) = \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S] = \mathbb{E} \left[ \left( \sum_{i=0}^n i^2 M_i^2 + \sum_{i \neq j} ij M_i M_j \right) \left( \sum_{i=0}^n M_i^2 + \sum_{i \neq j} M_i M_j \right) \right] - \mathbb{E}[R]\mathbb{E}[S].$$

The covariance term proves to be too complicated to calculate, because of the first term in the equation above.

Another way we can estimate the variance of the center of mass is by using first-order Taylor approximation from the same paper mentioned above, which is as follows

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

We will prove that

$$\text{SD}(CM) \approx \sqrt{\frac{1}{12} \frac{n(n+2)(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2}}. \quad (3.3.2)$$

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

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

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E} \left[ \sum_{i=0}^n M_i \right] = (n+1)\alpha_1, \\ \sigma_S^2 &= \text{Var} \left( \sum_{i=0}^n M_i \right) = \sum_{i=0}^n \text{Var}(M_i) = (n+1)(\alpha_2 - \alpha_1^2), \\ \sigma_R^2 &= \sum_{i=0}^n i^2 \text{Var}(M_i) = \frac{n(n+1)(2n+1)}{6}(\alpha_2 - \alpha_1^2).\end{aligned}$$

To find the covariance, we use the second-order Taylor approximation for expected value of the ratio of two variables to solve for covariance, and more precisely, the term closest to the one we need ( $2\frac{\text{Cov}(R,S)}{\mu_S}$ ):

$$\begin{aligned}\mathbb{E}(R/S) &\approx \frac{\mu_R}{\mu_S} - \frac{\text{Cov}(R,S)}{(\mu_S)^2} + \frac{\text{Var}(S)\mu_R}{(\mu_S)^3} \\ \frac{n}{2} &= \frac{n}{2} - \frac{\text{Cov}(R,S)}{(\mu_S)^2} + \frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\mu_S^3} \\ \frac{\text{Cov}(R,S)}{\mu_S} &= \frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\mu_S^2}.\end{aligned}$$

Now we calculate each term of our formula for variance:

$$\begin{aligned}\frac{\mu_R^2}{\mu_S^2} &= \frac{n^2}{4}, \\ \frac{\sigma_R^2}{\mu_R^2} &= \frac{\frac{n(n+1)(2n+1)}{6}(\alpha_2 - \alpha_1^2)}{\frac{n^2(n+1)^2}{4}\alpha_1^2} = \frac{\frac{1}{6}(2n+1)(\alpha_2 - \alpha_1^2)}{\frac{1}{4}n(n+1)\alpha_1^2}, \\ 2\frac{\text{Cov}(R,S)}{\mu_R\mu_S} &= 2\frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\mu_R\mu_S^2} = 2\frac{(n+1)(\alpha_2 - \alpha_1^2)\frac{n(n+1)}{2}\alpha_1}{\frac{n(n+1)}{2}\alpha_1(n+1)^2\alpha_1^2} = 2\frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2}, \\ \frac{\sigma_S^2}{\mu_S^2} &= \frac{(n+1)(\alpha_2 - \alpha_1^2)}{(n+1)^2\alpha_1^2} = \frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2}.\end{aligned}$$

Hence, we have

$$\begin{aligned}\text{Var}(R/S) &\approx \frac{n^2}{4} \left( \frac{\frac{1}{6}(2n+1)(\alpha_2 - \alpha_1^2)}{\frac{1}{4}n(n+1)\alpha_1^2} - 2\frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} + \frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} \right) = \\ &= \frac{n^2}{4} \left[ \frac{\frac{2}{3}(2n+1)(\alpha_2 - \alpha_1^2)}{n(n+1)\alpha_1^2} - \frac{(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} \right] = \frac{\frac{1}{6}n(2n+1)(\alpha_2 - \alpha_1^2)}{n(n+1)\alpha_1^2} - \frac{\frac{1}{4}(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} = \\ &= \frac{(\frac{1}{12}n^2 + \frac{1}{6}n)(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} = \frac{1}{12} \frac{(n^2 + 2n)(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} = \frac{1}{12} \frac{n(n+2)(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2}.\end{aligned}$$

Thus, the second approximation of the standard deviation, which we will call  $a_2$ , would be

$$\text{SD}(CM) \approx \sqrt{\frac{1}{12} \frac{n(n+2)(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2}}. \quad (a_2)$$

### 3.4 Simulations and Exact Value

In this section we will look at the results obtained from the simulations and ones we get from the exact value formula (Theorem 3.2.1.). I wrote a code in R to conduct simulations for the standard deviation in this configuration, using the provided definition of the center of mass,  $CM = \frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i}$ . The results we obtain by setting  $m = 1$ ,  $M = 10$  and varying probability  $p$  as well as the number of nodes  $n$  are portrayed in the table below:

<b>m = 1, M = 10</b>	<b>Standard Deviation from Simulations</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	0.82604	2.71196	8.43798
<b>0.2</b>	1.28927	4.02143	12.64450
<b>0.3</b>	1.77995	5.14317	16.40000
<b>0.4</b>	2.05300	6.28956	19.96200
<b>0.5</b>	2.39161	7.44943	23.52660
<b>0.6</b>	2.80841	8.54955	27.90520
<b>0.7</b>	3.23007	10.32340	31.58960
<b>0.8</b>	3.66563	12.25760	37.16310
<b>0.9</b>	4.08786	13.35720	40.16810

Figure 3.4.1. Standard deviation from simulations for the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M = 10$ ,  $p$  is the probability, and  $n$  is the number of nodes.

It appears that the standard deviation increases as  $n$  increases. For these specific parameters, the results also increase as the probability increases. The simulations also indicate that we obtain different results if we vary any of the variables - including the values of the masses.

<b>m = 1, M = 10</b>	<b>Standard Deviation from Exact Value Formula</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	0.86856	2.71237	8.56631
<b>0.2</b>	1.28568	4.01353	12.67538
<b>0.3</b>	1.65492	5.16508	16.31183
<b>0.4</b>	2.01840	6.29833	19.89035
<b>0.5</b>	2.39734	7.48018	23.62243
<b>0.6</b>	2.80788	8.76283	27.67351
<b>0.7</b>	3.26191	10.18970	32.18253
<b>0.8</b>	3.74838	11.74920	37.11927
<b>0.9</b>	4.08496	12.96880	41.02120

Figure 3.4.2. Standard deviation from exact value formula for the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M = 10$ ,  $p$  is the probability, and  $n$  is the number of nodes.

Since the results from simulations are prone to variation, we compute the formula for the exact value (3.2.2), stated in the Theorem 3.2.1 in section 2 of this chapter, as shown in the table above. The code for the simulations, as well as the exact value formula are included in Appendix A.1. Since R cannot compute the formula for big values of  $n$ , I also included the Mathematica code for the exact value formula which we used to calculate the standard deviation for big  $n$ 's in Appendix B.2. The table above presents standard deviation for  $m = 1$ ,  $M = 10$ , and varied  $p$  and  $n$  so that we can compare it with results from the simulations. These results are undoubtedly very close to those we obtained from the simulations, which indicates that the simulations are fairly accurate.

We notice that as we increase the value of mass  $M$ , there is a limit to which the standard deviation tends to. We will thus record the standard deviation obtained through maximizing the  $M$  value so that we get the maximum standard deviation we can get from the software. The table below summarizes results for standard deviation of the center of mass obtained through this sampling in Mathematica:

<b>m = 1, M -&gt; infinity</b>	<b>Standard Deviation from Exact Value Formula</b>				
<b>P</b>	<b>n = 100</b>	<b>n = 500</b>	<b>n = 1000</b>	<b>n = 5000</b>	<b>n = 10000</b>
<b>0.1</b>	0.97722	2.15835	3.04764	6.80626	9.62400
<b>0.2</b>	1.46687	3.23798	4.57177	10.20950	14.43610
<b>0.3</b>	1.92233	4.24027	5.98639	13.36764	18.90146
<b>0.4</b>	2.40050	5.28981	7.46723	16.67278	23.57455
<b>0.5</b>	2.94508	6.48084	9.14699	20.42058	28.87329
<b>0.6</b>	3.61647	7.94139	11.20550	25.01125	35.36330
<b>0.7</b>	4.53079	9.91302	13.98160	31.19710	44.10762
<b>0.8</b>	5.98814	13.00130	18.32170	40.85340	57.75525
<b>0.9</b>	9.28273	19.60400	27.55280	61.31090	86.65460

Figure 3.4.3. Exact value standard deviation for the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M$  is the other mass as the mass tends to  $+\infty$ ,  $p$  is the probability, and  $n$  is the number of nodes.

What's more, 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 values of masses. Intuitively, the standard deviations in the  $n = 100$  columns seem to differ

by around a factor of a little over 3 compared to those in column  $n = 1000$ . It seems to be the case in  $n = 1000$  and  $n = 10000$ . It is possible that since  $n$  increases by a factor of 10 in those cases, then the standard deviation increases by approximately  $\sqrt{10} \approx 3.16227766017$ . This will be further analyzed later in this section.

It is possible to create a table with results of standard deviation for a basic case, where  $m = 1$  and  $M = 0$ . In this case, the exact value formula (Theorem 3.2.1),

$$SD(CM) = \sqrt{-\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}},$$

would give us 0 in the denominator of the fraction  $\frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{km^2 + (n+1-k)M^2}$ . More specifically, we encounter this issue when, in case of  $m = 1$  and  $M = 0$ , we perform the summation for  $k = 0$ , so that in the denominator, we have

$$\frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{km^2 + (n+1-k)M^2} = \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{0 \cdot m^2 + (n+1-k)0^2} = \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{0}.$$

That said, if, say,  $k = 0$ , that means there are 0 of the masses  $m$ , and so the whole system only includes masses  $M = 0$ . This case can therefore be disregarded, because all the masses are 0, so the center of mass is not defined in the first place. Hence, the table with standard deviations for  $m = 1$  and  $M = 0$  is provided below:

<b>m = 1, M = 0</b>	<b>Standard Deviation from Exact Value Formula</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 500</b>	<b>n = 1000</b>
<b>0.1</b>	9.28272	19.60400	27.55280
<b>0.2</b>	5.98814	13.00130	18.32170
<b>0.3</b>	4.53079	9.91302	13.98160
<b>0.4</b>	3.61647	7.94139	11.20550
<b>0.5</b>	2.94509	6.48084	9.14699
<b>0.6</b>	2.40050	5.28981	7.46723
<b>0.7</b>	1.92233	4.24027	5.98639
<b>0.8</b>	1.46687	3.23798	4.57177
<b>0.9</b>	0.97722	2.15835	3.04764

Figure 3.4.4. Exact Value standard deviation for the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M = 0$  is the other mass,  $p$  is the probability, and  $n$  is the number of nodes.

Again, we can see that there is some relationship between standard deviations in columns with varying  $n$ 's. More specifically, if for the same values of  $p$ ,  $m$  and  $M$  we increase the number of nodes  $n$  by  $x$ , the standard deviation seems to increase by a factor of  $\sqrt{x}$ . Say, if we consider  $n = 100$  and  $n = 500$  columns, the  $n$  increases by a factor of 5, and at the same time, the standard deviation seems to increase by a little over 2, possibly the square root of 5. For instance, if we compare the standard deviations in this case in the row with probability  $p = 0.1$ , and  $n = 100$ ,  $n = 500$ , and look at the ratio of the latter to the former, we get

$$\frac{19.604}{9.28272} \approx 2.111881,$$

and the square root of 5 is

$$\sqrt{5} \approx 2.2360679775.$$

If we look at the same row, and  $n = 500$ ,  $n = 1000$ , where  $n$  increases by a factor of 2, we would get

$$\frac{27.5528}{19.604} \approx 1.405468272,$$

and the square root of 2 is

$$\sqrt{2} \approx 1.41421356237.$$

Although the exact value formula does not help in explaining this behavior of the standard deviation, both of the approximations shed some light on to why the patterns above occur. The approximations we have are:

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(\alpha_2 + n\alpha_1^2)}} \quad (a_1)$$

and

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)(\alpha_2 - \alpha_1^2)}{12(n+1)\alpha_1^2}}. \quad (a_2)$$

We notice that  $\alpha_1 = mp + M(1-p)$  and  $\alpha_2 = m^2p + M^2(1-p)$  both depend on the probability  $p$ , and the value of masses  $m$  and  $M$ , and not the number of nodes  $n$ . Thus, if we fix the values of these variables, we can rewrite the 2nd approximation as follows:

$$\text{SD}(CM) \approx \sqrt{\frac{n(n+2)}{n+1} \cdot \frac{\alpha_2 - \alpha_1^2}{12\alpha_1^2}} = \sqrt{\frac{n(n+2)}{n+1}} \cdot A$$

where

$$A = \sqrt{\frac{\alpha_2 - \alpha_1^2}{12\alpha_1^2}}$$

is some constant given values of  $p$ ,  $m$  and  $M$ . If we increase  $n$  by a factor of  $x$ , we would have

$$\text{SD}(CM) \approx \sqrt{\frac{xn(xn+2)}{xn+1}} \cdot A,$$

meaning the standard deviation would increase by

$$\frac{\sqrt{\frac{xn(xn+2)}{xn+1}} \cdot A}{\sqrt{\frac{n(n+2)}{n+1}} \cdot A} = \frac{\sqrt{\frac{xn(xn+2)}{xn+1}}}{\sqrt{\frac{n(n+2)}{n+1}}} = \sqrt{\frac{xn(xn+2)}{xn+1} \cdot \frac{n+1}{n(n+2)}} = \sqrt{\frac{x(xn+2)}{xn+1} \cdot \frac{n+1}{n+2}},$$

which is close to  $\sqrt{x}$  as  $n$  goes to infinity in which cause we could cancel  $xn+2$  with  $xn+1$ , and  $n+1$  with  $n+2$  to get an approximation. That is why as we have seen in this section, if we increase  $n$  from 100 to 5000 (by a factor of 5), then the standard deviation increases by:

$$\sqrt{\frac{5 \cdot 100(5 \cdot 100 + 2)}{5 \cdot 100 + 1} \cdot \frac{100 + 1}{100(100 + 2)}} \approx 2.22729939356,$$

whereas

$$\sqrt{5} \approx 2.2360679775.$$

We can come to the same conclusion by considering the first approximation, although it is not as clear as in the second approximation. We can see that since the dominating term in the numerator is  $n(n+2) \approx n^2$  and since the denominator has one  $n$ , the standard deviation would change by roughly the square root of whatever  $n$  is increased by. Since we are using approximations for analysis, the increase in the standard deviation is an approximations as well.

Another thing worth considering is that if we look at the tables with  $m = 1$ ,  $M \rightarrow \infty$  and  $m = 1$ ,  $M = 0$ , we notice that the columns with results for  $n = 100$ ,  $n = 500$ ,  $n = 1000$  in both of the tables are flipped versions of each other. This suggests that, more precisely, the standard deviation of the center of mass in this case depends on the ratio of the masses  $m/M$  (as opposed to the value of each of the masses separately), in addition to probability  $p$  and number of nodes  $n$ . This is because for  $m = 1$ ,  $M \rightarrow \infty$ , the ratio goes to 0, and for  $m = 1$ ,  $M \rightarrow 0$ , the ratio goes to infinity. Below are two more tables, which provide results for  $m = 1$ ,  $M = 1000$  (ratio  $m/M = 0.001$ ) and  $m = 1$ ,  $M = 0.001$  (ratio  $m/M = 1000$ ):



<b>m = 1, M = 1000</b>	<b>Standard Deviation from Exact Value Formula</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 500</b>	<b>n = 1000</b>
<b>0.1</b>	0.97602	2.15595	3.04425
<b>0.2</b>	1.46501	3.23392	4.56605
<b>0.3</b>	1.91953	4.23419	5.97783
<b>0.4</b>	2.39641	5.28096	7.45477
<b>0.5</b>	2.93902	6.46782	9.12865
<b>0.6</b>	3.60706	7.92141	11.17750
<b>0.7</b>	4.51479	9.87968	13.93490
<b>0.8</b>	5.95516	12.93540	18.22960
<b>0.9</b>	9.16254	19.40200	27.27440

Figure 3.4.5. Exact value standard deviation for the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M = 1000$  is the other mass,  $p$  is the probability, and  $n$  is the number of nodes.

<b>m = 1, M = 0.001</b>	<b>Standard Deviation from Exact Value Formula</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 500</b>	<b>n = 1000</b>
<b>0.1</b>	9.16254	19.40200	27.27440
<b>0.2</b>	5.95516	12.93540	18.22960
<b>0.3</b>	4.51479	9.87968	13.93490
<b>0.4</b>	3.60706	7.92141	11.17750
<b>0.5</b>	2.93902	6.46782	9.12865
<b>0.6</b>	2.39641	5.28096	7.45477
<b>0.7</b>	1.91953	4.23419	5.97783
<b>0.8</b>	1.46501	3.23392	4.56605
<b>0.9</b>	0.97602	2.15595	3.04425

Figure 3.4.6. Exact value standard deviation for the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M = 0.001$  is the other mass,  $p$  is the probability, and  $n$  is the number of nodes.

Looking at the exact value formula,

$$\text{SD}(CM) = \sqrt{-\frac{n+2}{12} + \frac{(n+1)(n+2)}{12} \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}},$$

we can tell that these observations are accurate - considering the term  $\frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}$ , if we multiply both  $m$  and  $M$  by some constant  $c$  (so that their ratio stays the same), the value of the term does not change:

$$\frac{k(cm)^2 + (n+1-k)(cM)^2}{(k(cm) + (n+1-k)(cM))^2} = \frac{c^2(km^2 + (n+1-k)M^2)}{c^2(km + (n+1-k)M)^2} = \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}.$$

### 3.5 Analysis of Results

In this section, we will analyze all the results we obtained for the standard deviation of the center of mass in the one-dimensional discrete case in depth. We compare the approximations for the standard deviation by calculating relative errors with respect to the numbers we get from the exact value formula stated in Theorem 3.2.1. ( $\sigma$ ). We create tables with 8 columns: the 1st column from the left provides information on the parameters used, the 2nd column provides results from simulations denoted as  $s$  (an average of 10 random samples to lessen the variability), the 3rd column provides the results from the exact value formula and is denoted as  $\sigma$ . The 4th and 5th columns compute approximations from formulas 3.3.1 (denoted as  $a_1$ ) and 3.3.2 (denoted as  $a_2$ ) respectively. The 6th and 7th columns compare the approximations  $a_1$  and  $a_2$  to  $\sigma$  by computing the relative error. The last column (on the very right) depicts the ratio  $|\sigma - a_1|/|\sigma - a_2|$ , i.e., it compares which approximation gives results closer to the actual value ( $\sigma$ ). If the ratio is greater than 1, then the second approximation ( $a_2$ ) is closer to the standard deviation obtained through the exact value formula ( $\sigma$ ), and if the ratio is less than 1, then the first approximation ( $a_1$ ) yields values closer to the actual value. Below are two of the tables summarizing the data with results for  $m = 1$ , varied  $M$ 's,  $p = 0.5$ , and  $n = 100$ ,  $n = 1000$ ,  $n = 10000$ .

M	Standard Deviation				Relative Error in Approx.		Comparison
	Simulation (s)	Exact Value ( $\sigma$ )	Approx. 3.3.1 (a1)	Approx. 3.3.2 (a2)	$ \sigma - a_1  / a_1$	$ \sigma - a_2  / a_2$	
0.001	2.88417	2.93902	2.89521	2.88104	0.01513	0.02012	0.75561
0.01	2.91958	2.88500	2.84356	2.83013	0.01457	0.01939	0.75524
0.1	2.42947	2.39734	2.37355	2.36572	0.01002	0.01337	0.75237
1	0.00000	0.00000	0.00000	0.00000	-	-	-
10	2.40073	2.39734	2.37355	2.36572	0.01002	0.01337	0.75237
100	2.90470	2.88500	2.84356	2.83013	0.01457	0.01939	0.75524
1000	2.92318	2.93902	2.89521	2.88104	0.01513	0.02012	0.75561

Figure 3.5.1. Standard deviation of the center of mass in the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M$  is the other mass,  $p = 0.5$  is the probability, and  $n = 100$  is the number of nodes. The standard deviation from simulations is denoted by  $(s)$ , the exact value is denoted by  $\sigma$ , the approximation 3.3.1 is denoted by  $a_1$ , and the approximation 3.3.2, is denoted by  $a_2$ .

M	Standard Deviation				Relative Error in Approx.		Comparison
	Simulation (s)	Exact Value ( $\sigma$ )	Approx. 3.3.1 (a1)	Approx. 3.3.2 (a2)	$ \sigma - a1  / a1$	$ \sigma - a2  / a2$	
0.001	9.01196	9.12865	9.11502	9.11049	0.00150	0.00199	0.75055
0.01	9.01425	8.96533	8.95241	8.94812	0.00144	0.00192	0.75073
0.1	7.36740	7.48018	7.47267	7.47018	0.00100	0.00134	0.75100
1	0.00000	0.00000	0.00000	0.00000	-	-	-
10	7.52343	7.48018	7.47267	7.47018	0.00100	0.00134	0.75100
100	9.04523	8.96533	8.95241	8.94812	0.00144	0.00192	0.75073
1000	8.99009	9.12865	9.11502	9.11049	0.00150	0.00199	0.75055

Figure 3.5.2. Standard deviation of the center of mass in the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M$  is the other mass,  $p = 0.5$  is the probability, and  $n = 1000$  is the number of nodes.

M	Standard Deviation				Relative Error in Approx.		Comparison
	Simulation (s)	Exact Value ( $\sigma$ )	Approx. 3.3.1 (a1)	Approx. 3.3.2 (a2)	$ \sigma - a1  / a1$	$ \sigma - a2  / a2$	
0.001	28.16360	28.81558	28.81130	28.80980	0.00015	0.00020	0.74055
0.01	28.03448	28.30137	28.29730	28.29590	0.00014	0.00019	0.74419
0.1	23.81158	23.62243	23.62010	23.61930	0.00010	0.00013	0.74418
1	0.00000	0.00000	0.00000	0.00000	-	-	-
10	23.49276	23.62243	23.62010	23.61930	0.00010	0.00013	0.74418
100	28.18486	28.30137	28.29730	28.29590	0.00014	0.00019	0.74419
1000	28.14486	28.81558	28.81130	28.80980	0.00015	0.00020	0.74055

Figure 3.5.3. Standard deviation of the center of mass in the one-dimensional discrete case, where  $m = 1$  is one of the masses,  $M$  is the other mass,  $p = 0.5$  is the probability, and  $n = 10000$  is the number of nodes.

Again, for  $n = 10000$ , we have to use Mathematica to compute the standard deviation from the exact value formula as it contains a summation for large  $n$ , which cannot be computed in RStudio. The relative errors we see in the tables indicate that both of the approximations are very close to the results from the exact value formula. The last column from the left, which compares the approximations by computing the ratio of their differences as compared to  $\sigma$ , tells us that the approximation  $a_2$  which uses first-order Taylor approximation for variance is a better estimate for the standard deviation of the center of mass.

Last but not least, as part of our analysis of results, let us go back to the results obtained from the exact value formula in the previous section. It is worth mentioning that if we were to standardize our results (as shown below, 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, in chapter 4 section 6 when we compare it with the standard deviation of the one-dimensional uniform case. The standardization is as follows

$$CM = \frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i} \xrightarrow{\text{standardize}} CM = \frac{\sum_{i=0}^n (i/n)M_i}{\sum_{i=0}^n M_i} = \frac{1}{n} \cdot \frac{\sum_{i=0}^n iM_i}{\sum_{i=0}^n M_i}.$$

Standardizing would cause the positions of the masses to be within the  $(0, 1)$  interval.



# 4

## Center of Mass in the One-dimensional Uniform Case

### 4.1 Introduction to the Case and Simulations

Let us consider 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

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

where  $Y_i$  is a uniform random variable, which uniformly assigns a position between 0 and 1 to the  $i$ th mass,  $M_i$ . The variables  $M_i$  and  $Y_i$  are independent of each other, and calculating the expected value and standard deviation of the former is not a problem. With  $Y_i$  coming from a uniform distribution between 0 and 1, it follows that

$$\mathbb{E}[Y_i] = \frac{0 + 1}{2} = \frac{1}{2},$$

and

$$\mathbb{E}[Y_i^2] = \int_0^1 u^2 f_U(u) du = \int_0^1 u^2 \cdot 1 du = \frac{1}{3} u^3 \Big|_{u=0}^{u=1} = \frac{1}{3}.$$

We modify the code of the discrete case so that the positions are assigned uniformly. The code is provided in Appendix A.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 section we will find the expected value of the center of mass in this uniform case, and after that we will look at variance and standard deviation.

## 4.2 Expected Value

We expect the expected value of the center of mass to be  $\frac{1}{2}$ , since the positions are assigned randomly from the uniform function, between 0 and 1. We will provide a rigorous proof for the finding, as well as approximations that indicate similar value.

**Theorem 4.2.1.** *The expected value of the center of mass in the one-dimensional uniform case is*

$$\mathbb{E}[CM] = \frac{1}{2}. \quad (4.2.1)$$

**Proof.** To prove this, we will use total expectation theorem and conditional expectation definition, where  $M_{\text{tot}}$  is the total mass of the configuration,  $k$  is the total number of masses  $m$  in the configuration, which occur with probability  $p$ , and  $n + 1 - k$  is the number of masses  $M$ , which occur with probability  $1 - p$ :

$$\mathbb{E}[CM] = \sum \mathbb{E}[CM \mid \text{Mass} = M_{\text{tot}}] \mathbb{P}(\text{Mass} = M_{\text{tot}}).$$

We also have that

$$\mathbb{P}(\text{Mass} = M_{\text{tot}}) = \binom{n+1}{k} p^k (1-p)^{n+1-k}$$

and

$$M_{\text{tot}} = \sum_{i=0}^n M_i = km + (n+1-k)M.$$

We can now calculate the expected value as follows:

$$\mathbb{E}[CM] = \mathbb{E} \left[ \frac{\sum Y_i M_i}{\sum M_i} \right] = \sum \mathbb{E} \left[ \frac{\sum Y_i M_i}{\sum M_i} \middle| \text{Mass} = M_{\text{tot}} \right] \mathbb{P}(\text{Mass} = M_{\text{tot}}).$$

Since  $Y_i$  and  $M_i$ , as well as  $\sum_{i=0}^n Y_i M_i$  and  $\sum_{i=0}^n M_i$  are independent variables (because of the given total mass of the system), and since  $\mathbb{E}[Y_i] = \frac{1}{2}$ , we have

$$\mathbb{E} \left[ \sum Y_i M_i \mid M_{\text{tot}} \right] = \sum_{i=0}^n \mathbb{E}[Y_i M_i \mid M_{\text{tot}}] = \sum_{i=0}^n \mathbb{E}[Y_i \mid M_{\text{tot}}] \mathbb{E}[M_i \mid M_{\text{tot}}] = \frac{1}{2} \sum_{i=0}^n \mathbb{E}[M_i \mid M_{\text{tot}}] =$$

$$= \frac{1}{2}(n+1) \cdot \frac{km + (n+1-k)M}{n+1} = \frac{1}{2}km + (n+1-k)M.$$

Thus, we have

$$\begin{aligned} \mathbb{E}[CM] &= \sum \mathbb{E} \left[ \frac{\sum Y_i M_i}{\sum M_i} \middle| \text{Mass} = M_{\text{tot}} \right] \mathbb{P}(\text{Mass} = M_{\text{tot}}) = \\ &= \sum \frac{\mathbb{E}[\sum Y_i M_i \mid \text{Mass} = M_{\text{tot}}]}{\mathbb{E}[\sum M_i \mid \text{Mass} = M_{\text{tot}}]} \mathbb{P}(\text{Mass} = M_{\text{tot}}) = \sum \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{km + (n+1-k)M} \cdot \frac{1}{2} [km + (n+1-k)M] = \\ &= \frac{1}{2} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} = \frac{1}{2}. \end{aligned}$$

□

As for the approximations for this case, we can try to use first-order Taylor approximation.

We have that

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

where  $R = \sum_{i=0}^n Y_i M_i$  and  $S = \sum_{i=0}^n M_i$ . 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  $M_i$  and  $Y_i$  are independent, it follows that  $\mathbb{E}[M_i Y_i] = \mathbb{E}[M_i] \mathbb{E}[Y_i]$ . Thus, we can calculate needed components:

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

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

$$\begin{aligned} \sigma_R^2 &= \text{Var} \left( \sum_{i=0}^n Y_i M_i \right) = \sum_{i=0}^n \text{Var}(Y_i M_i) = \sum_{i=0}^n [\mathbb{E}[M_i]^2 \text{Var}(Y_i) + \mathbb{E}[Y_i]^2 \text{Var}(M_i) + \text{Var}(Y_i) \text{Var}(M_i)] = \\ &= \sum_{i=0}^n \left[ \alpha_1^2 \frac{1}{12} + \frac{1}{4}(\alpha_2 - \alpha_1^2) + \frac{1}{12}(\alpha_2 - \alpha_1^2) \right] = \sum_{i=0}^n \left[ \frac{1}{12} \alpha_1^2 + \frac{1}{4} \alpha_2 - \frac{1}{4} \alpha_1^2 + \frac{1}{12} \alpha_2 - \frac{1}{12} \alpha_1^2 \right] = \\ &= \sum_{i=0}^n \left[ \frac{1}{3} \alpha_2 - \frac{1}{4} \alpha_1^2 \right] = (n+1) \left( \frac{1}{3} \alpha_2 - \frac{1}{4} \alpha_1^2 \right), \end{aligned}$$

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

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



$$\begin{aligned}
&= \sum_{i=0}^n \mathbb{E}[Y_i] \mathbb{E}[M_i^2] + \sum_{i \neq j}^n \mathbb{E}[Y_i] \mathbb{E}[M_i M_j] - \mu_R \mu_S = \frac{1}{2}(n+1)\alpha_2 + \frac{1}{2} \cdot 2 \cdot \frac{n(n+1)}{2} \alpha_1^2 - \frac{1}{2}(n+1)^2 \alpha_1^2 = \\
&= \frac{1}{2}(n+1) [\alpha_2 + n\alpha_1 - (n+1)\alpha_1] = \frac{1}{2}(n+1) [\alpha_2 - \alpha_1^2].
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}(R/S) &\approx \frac{\frac{1}{2}(n+1)\alpha_1}{(n+1)\alpha_1} - \frac{\frac{1}{2}(n+1)[\alpha_2 - \alpha_1^2]}{(n+1)^2 \alpha_1^2} + \frac{(n+1)(\alpha_2 - \alpha_1^2) \frac{1}{2}(n+1)\alpha_1}{(n+1)^3 \alpha_1^2} = \\
&= \frac{1}{2} - \frac{\alpha_2 - \alpha_1^2}{2(n+1)\alpha_1} + \frac{\alpha_2 - \alpha_1^2}{2(n+1)\alpha_1} = \frac{1}{2} - \left( \frac{\alpha_2 - \alpha_1^2 - \alpha_2 + \alpha_1^2}{2(n+1)\alpha_1} \right) = \frac{1}{2} - \left( \frac{\alpha_2 - \alpha_1^2 - \alpha_2 + \alpha_1^2}{2(n+1)\alpha_1} \right) = \frac{1}{2}.
\end{aligned}$$

Therefore, by second-order Taylor approximation for expected value of the ratio of two random variables, the estimated expected value of the center of mass in this uniform case is

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

which is what we expected it to be.

Concluding, both the rigorous proof with the given mass, as well as the approximations confirm that the expected value of the center of mass, where positions are assigned uniformly between 0 and 1, is equal to  $\frac{1}{2}$ . We will also see later on that the simulations turn out to be accurate since the expected value of the center of mass fluctuates around the value of  $\frac{1}{2}$ .

### 4.3 Formula for Variance and Standard Deviation

In this section, we will derive a formula for variance, and subsequently for standard deviation of the center of mass. We can find the standard deviation of the center of mass in this one-dimensional uniform case by first calculating the variance from formula

$$\text{Var}(CM) = \mathbb{E}[CM^2] - \mathbb{E}[CM]^2.$$

In the previous section, we found that  $\mathbb{E}[CM] = \frac{1}{2}$  (Theorem 4.2.1), thus we need to compute the first term of the formula above. This section will prove the following theorem:

**Theorem 4.3.1.** *The variance and standard deviation of the center of mass in the one-dimensional uniform case are*

$$\text{Var}(CM) = \frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2} \quad (4.3.1)$$

and

$$\text{SD}(CM) = \sqrt{\frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}}. \quad (4.3.2)$$

**Proof.** We can prove this theorem in the same way we proved the theorem for the standard deviation of the center of mass in the one-dimensional discrete case. The only thing that differs is the formula for the center of mass, since in this case the positions are chosen uniformly. Let  $CM$  be discrete random variable, and let the total mass variable, say,  $\text{Mass} = M_{\text{tot}}$ . According to the theorem,  $\mathbb{E}[CM^2]$  is equal to the expected value of the conditional expected value of  $CM^2$  given that  $\text{Mass} = M_{\text{tot}}$ . In other words, if we let  $M_{\text{tot}}$  be the total mass of the configuration,  $M_i$  be a random variable that assigns a mass  $m$  with probability  $p$  and a mass  $M$  with probability  $1-p$ , and  $i$  be the position of the mass  $M_i$  on this one-dimensional lattice, then we would have that we would need to find the following:

$$\mathbb{E}[CM^2] = \sum \mathbb{E}[CM^2 \mid \text{Mass} = M_{\text{tot}}] \mathbb{P}(\text{Mass} = M_{\text{tot}}).$$

Let  $k$  be the total number of masses  $m$  in the configuration, and  $n+1-k$  be the number of masses  $M$ . We have that:

$$\begin{aligned} \mathbb{E}[CM^2] &= \sum \mathbb{E} \left[ \frac{(\sum Y_i M_i)^2}{(\sum M_i)^2} \middle| \text{Mass} = M_{\text{tot}} \right] \mathbb{P}(\text{Mass} = M_{\text{tot}}) = \\ &= \sum \frac{\mathbb{E}[(\sum Y_i M_i)^2 \mid M_{\text{tot}}]}{\mathbb{E}[(\sum M_i)^2 \mid M_{\text{tot}}]} \mathbb{P}(\text{Mass} = M_{\text{tot}}) = \sum \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n+1-k)M)^2} \mathbb{E} \left[ \left( \sum Y_i M_i \right)^2 \middle| M_{\text{tot}} \right]. \end{aligned}$$

Now we need to calculate the tricky part of this equation, which is the latter term. Since  $Y_i$  and  $M_i$  are independent, we have:

$$\mathbb{E} \left[ \left( \sum Y_i M_i \right)^2 \middle| \text{Mass} = M_{\text{tot}} \right] = \mathbb{E} \left[ \sum_{i=0}^n Y_i^2 M_i^2 + \sum_{i \neq j} Y_i Y_j M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] =$$

$$\begin{aligned}
&= \mathbb{E} [Y_i^2] \mathbb{E} \left[ \sum_{i=0}^n M_i^2 \right] + \mathbb{E}[Y_i] \mathbb{E}[Y_j] \mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = \\
&= \frac{1}{3} \sum_{i=0}^n \mathbb{E} [M_i^2] + \frac{1}{2} \cdot \frac{1}{2} \mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = \\
&= \frac{1}{3} (n+1) \frac{km^2 + (n+1-k)M^2}{n+1} + \frac{1}{4} \mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = \\
&= \frac{1}{3} km^2 + (n+1-k)M^2 + \frac{1}{4} \mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right].
\end{aligned}$$

It turns out that  $\mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = (km + (n+1-k)M)^2 - (km^2 + (n+1-k)M^2)$ .

This is because, if we write out the summation term in order, for  $i = 0$  to  $i = n$ , we get

$$\begin{aligned}
&M_{\text{tot}}M_1 + M_{\text{tot}}M_1 + M_{\text{tot}}M_2 + \cdots + M_{\text{tot}}M_n \\
&M_1M_{\text{tot}} + M_1M_2 + M_1M_3 + \cdots + M_1M_n \\
&\vdots \\
&M_nM_{\text{tot}} + M_nM_1 + M_nM_2 + \cdots + M_nM_{n-1}.
\end{aligned}$$

We notice that we can add terms where  $i = j$  and subtract them in the following way:

$$\begin{aligned}
&M_{\text{tot}}M_1 + M_{\text{tot}}M_1 + M_{\text{tot}}M_2 + \cdots + M_{\text{tot}}M_n (+M_{\text{tot}}M_{\text{tot}} - M_{\text{tot}}M_{\text{tot}}) \\
&M_1M_{\text{tot}} + M_1M_2 + M_1M_3 + \cdots + M_1M_n (+M_1M_1 - M_1M_1) \\
&\vdots \\
&M_nM_{\text{tot}} + M_nM_1 + M_nM_2 + \cdots + M_nM_{n-1} (+M_nM_n - M_nM_n).
\end{aligned}$$

As a result, we can organize the terms as

$$\begin{aligned}
&M_{\text{tot}}M_{\text{tot}} + M_{\text{tot}}M_1 + M_1M_2 + M_{\text{tot}}M_3 + \cdots + M_{\text{tot}}M_n (-M_{\text{tot}}M_{\text{tot}}) \\
&M_1M_{\text{tot}} + M_1M_1 + M_1M_2 + M_1M_3 + \cdots + M_1M_n (-M_1M_1) \\
&\vdots
\end{aligned}$$

$$M_n M_{\text{tot}} + M_n M_1 + M_n M_2 + \cdots + M_n M_{n-1} + M_n M_n (-M_n M_n).$$

We then factor out the common term in each row as follows:

$$\begin{aligned} & M_{\text{tot}} (M_{\text{tot}} + M_1 + M_2 + M_3 + \cdots + M_n) - M_{\text{tot}} M_{\text{tot}} \\ & M_1 (M_{\text{tot}} + M_1 + M_2 + M_3 \cdots + M_n) - M_1 M_1 \\ & \quad \vdots \\ & M_n (M_{\text{tot}} + M_1 + M_2 + \dots M_{n-1} + M_n) - M_n M_n. \end{aligned}$$

We know that  $M_{\text{tot}} + M_1 + M_2 + M_3 + \cdots + M_n = \sum_{i=0}^n M_i = km + (n+1-k)M$ , and  $M_{\text{tot}} M_{\text{tot}} + M_1 M_1 + \cdots + M_n M_n = \sum_{i=0}^n M_i^2$ . We now sum up all of the terms above, and obtain

$$\begin{aligned} & M_{\text{tot}} \sum_{i=0}^n M_i + M_1 \sum_{i=0}^n M_i + \cdots + M_n \sum_{i=0}^n M_i - (M_{\text{tot}} M_{\text{tot}} + M_1 M_1 + \cdots + M_n M_n) = \\ & = \sum_{i=0}^n M_i (M_{\text{tot}} + M_1 + \cdots + M_n) - \sum_{i=0}^n M_i^2 = \left( \sum_{i=0}^n M_i \right)^2 - \sum_{i=0}^n M_i^2. \end{aligned}$$

Since  $\sum_{i=0}^n M_i = km + (n+1-k)M$ , which is a constant, we have that

$$\begin{aligned} & \mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = \mathbb{E} \left[ \left( \sum_{i=0}^n M_i \right)^2 - \sum_{i=0}^n M_i^2 \middle| \text{Mass} = M_{\text{tot}} \right] = \\ & = \mathbb{E} \left[ \left( \sum_{i=0}^n M_i \right)^2 \middle| \text{Mass} = M_{\text{tot}} \right] - \mathbb{E} \left[ \sum_{i=0}^n M_i^2 \middle| \text{Mass} = M_{\text{tot}} \right] = \\ & = \mathbb{E} \left[ \left( \sum_{i=0}^n M_i \right)^2 \middle| \text{Mass} = M_{\text{tot}} \right] - \sum_{i=0}^n \mathbb{E} \left[ M_i^2 \middle| \text{Mass} = M_{\text{tot}} \right] = \\ & = (km + (n+1-k)M)^2 - (n+1) \cdot \frac{km^2 + (n+1-k)M^2}{n+1} = (km + (n+1-k)M)^2 - (km^2 + (n+1-k)M^2). \end{aligned}$$

Thus,  $\mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = (km + (n+1-k)M)^2 - (km^2 + (n+1-k)M^2)$ . Hence, plugging in to the equation for  $\mathbb{E} \left[ \left( \sum Y_i M_i \right)^2 \middle| \text{Mass} = M_{\text{tot}} \right]$ , we have that

$$\begin{aligned} & \mathbb{E} \left[ \left( \sum Y_i M_i \right)^2 \middle| \text{Mass} = M_{\text{tot}} \right] = \frac{1}{3} km^2 + (n+1-k)M^2 + \frac{1}{4} \mathbb{E} \left[ \sum_{i \neq j} M_i M_j \middle| \text{Mass} = M_{\text{tot}} \right] = \\ & = \frac{1}{3} km^2 + (n+1-k)M^2 + \frac{1}{4} [(km + (n+1-k)M)^2 - (km^2 + (n+1-k)M^2)] = \end{aligned}$$

$$= \frac{1}{4}(km + (n + 1 - k)M)^2 + \frac{1}{12}(km^2 + (n + 1 - k)M^2).$$

Returning to our original equation, we have

$$\begin{aligned} \mathbb{E}[CM^2] &= \sum \frac{\binom{n+1}{k} p^k (1-p)^{n+1-k}}{(km + (n + 1 - k)M)^2} \left[ \frac{1}{4}(km + (n + 1 - k)M)^2 + \frac{1}{12}(km^2 + (n + 1 - k)M^2) \right] = \\ &= \frac{1}{4} + \frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n + 1 - k)M^2}{(km + (n + 1 - k)M)^2}. \end{aligned}$$

As a result, the formula for variance is as follows:

$$\begin{aligned} \text{Var}(CM) &= \mathbb{E}[CM^2] - \mathbb{E}[CM]^2 = \frac{1}{4} + \frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n + 1 - k)M^2}{(km + (n + 1 - k)M)^2} - \left(\frac{1}{2}\right)^2 = \\ &= \frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \frac{km^2 + (n + 1 - k)M^2}{(km + (n + 1 - k)M)^2}. \end{aligned}$$

□

We can use both RStudio and Mathematica to compute this summation, because R cannot compute it for  $n > 10^3 + 28$ . It appears to be the case that as  $n$  approaches infinity, the standard deviation goes to 0. One way we can justify this is by considering  $\binom{n+1}{k} p^k (1-p)^{n+1-k}$ , in which the binomial has the highest value at  $k = \frac{n+1}{2}$  for odd  $n$ 's. We would want to find a way to show that this binomial goes to 0 as  $n$  approaches infinity, so that the whole term goes to 0.

**Theorem 4.3.2.** *If  $p = \frac{1}{2}$ , then the standard deviation of the center of mass in the one-dimensional uniform case goes to 0 as  $n$  approaches infinity. In other words,*

$$\lim_{n \rightarrow \infty} SD(CM) = \lim_{n \rightarrow \infty} \sqrt{\frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} \left(\frac{1}{2}\right)^{n+1} \frac{km^2 + (n + 1 - k)M^2}{(km + (n + 1 - k)M)^2}} = 0$$

**Proof.** If we assume  $p = \frac{1}{2}$ , then for the highest possible value of the binomial we would have

$$\binom{n+1}{k} p^k (1-p)^{n+1-k} = \binom{n+1}{\frac{n+1}{2}} \left(\frac{1}{2}\right)^{n+1}.$$

The binomial coefficient can be estimated by Stirling's Approximation, which states that

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

where the sign  $\sim$  indicates that the two quantities are asymptotic, that is, their ratio goes to 1 when  $n$  approaches infinity. Applying this formula to our case, we have that

$$\begin{aligned} \binom{n+1}{\frac{n+1}{2}} &= \frac{(n+1)!}{\left(\frac{n+1}{2}\right)! \left(\frac{n+1}{2}\right)!} \sim \frac{\sqrt{2\pi(n+1)} \left(\frac{n+1}{e}\right)^{n+1}}{\left(\sqrt{2\pi \frac{n+1}{2}} \left(\frac{n+1}{e}\right)^{\frac{n+1}{2}}\right)^2} = \frac{\sqrt{2\pi(n+1)} \left(\frac{n+1}{e}\right)^{n+1}}{\frac{1}{2} \left(\sqrt{2\pi(n+1)}\right)^2 \left(\frac{n+1}{e}\right)^{n+1}} = \\ &= \frac{\left(\frac{n+1}{e}\right)^{n+1}}{\frac{1}{2}\sqrt{2\pi(n+1)}} \cdot \left(\frac{n+1}{e}\right)^{-(n+1)} = \frac{2^{n+1} \left(\frac{n+1}{e}\right)^{n+1}}{\frac{1}{2}\sqrt{2\pi(n+1)}} \cdot \left(\frac{n+1}{e}\right)^{-(n+1)} = \frac{2^{n+1}}{\sqrt{\frac{\pi}{2}(n+1)}}. \end{aligned}$$

Thus, it follows that for  $n$  approaching infinity,

$$\binom{n+1}{\frac{n+1}{2}} \left(\frac{1}{2}\right)^{n+1} \sim \frac{2^{n+1}}{\sqrt{\frac{\pi}{2}(n+1)}} \cdot \left(\frac{1}{2}\right)^{n+1} = \frac{1}{\sqrt{\frac{\pi}{2}(n+1)}} \rightarrow 0.$$

In other words, the biggest binomial coefficient for  $p = \frac{1}{2}$  goes to 0 as  $n$  approaches infinity. Since there are  $n+1$  terms in the summation in the standard deviation formula, that means that for  $p = \frac{1}{2}$  and for all  $n$ ,

$$\begin{aligned} &\frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} \left(\frac{1}{2}\right)^{n+1} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2} \leq \\ &\leq \frac{1}{12} (n+1) \frac{2^{n+1}}{\sqrt{\frac{\pi}{2}(n+1)}} \left(\frac{1}{2}\right)^{n+1} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}. \end{aligned}$$

Because

$$\lim_{n \rightarrow \infty} \frac{1}{12} (n+1) \frac{2^{n+1}}{\sqrt{\frac{\pi}{2}(n+1)}} \left(\frac{1}{2}\right)^{n+1} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2} = 0,$$

by the Squeeze Theorem:

$$\lim_{n \rightarrow \infty} \frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} \left(\frac{1}{2}\right)^{n+1} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2} = 0.$$

Hence, for  $p = \frac{1}{2}$ ,

$$\lim_{n \rightarrow \infty} \text{SD}(CM) = \lim_{n \rightarrow \infty} \sqrt{\frac{1}{12} \sum_{k=0}^{n+1} \binom{n+1}{k} \left(\frac{1}{2}\right)^{n+1} \frac{km^2 + (n+1-k)M^2}{(km + (n+1-k)M)^2}} = 0.$$

□

We can also try to simplify the summation in Mathematica hoping that we could prove the term goes to 0 as  $n$  goes to infinity, but just like in the one-dimensional discrete case, the result

contains hypergeometric functions. The input

$$\text{FullSimplify} \left[ \sum_{k=0}^{n+1} \frac{((n+1)(n+2)) (p^k \binom{n+1}{k} (1-p)^{-k+n+1}) (km^2 + M^2(-k+n+1))}{12(km + M(-k+n+1))^2} \right]$$

returns the following output:

$$\begin{aligned} & \frac{1}{12}(n+2)(1-p)^n \Gamma \left( \frac{m+Mn}{m-M} \right)^2 \\ & \left[ \frac{(n+1)^2 p(m+M) {}_3\tilde{F}_2 \left( -n, \frac{m+Mn}{m-M}, \frac{m+Mn}{m-M}, \frac{2m+M(n-1)}{m-M}, \frac{2m+M(n-1)}{m-M}, \frac{p}{p-1} \right)}{m-M} \right. \\ & \left. - (p-1) {}_3\tilde{F}_2 \left( -n-1, \frac{M(n+1)}{m-M}, \frac{M(n+1)}{m-M}, \frac{m+Mn}{m-M}, \frac{m+Mn}{m-M}, \frac{p}{p-1} \right) \right], \end{aligned}$$

where  ${}_3\tilde{F}_2(a_1, a_2, a_3; b_1, b_2; z)$  is a generalized hypergeometric function. The details on the function were provided at the end of section 3.2.

With that said, the next section will derive two approximations for the standard deviation, both of which can be proven to go to 0 as  $n$  goes to infinity.

#### 4.4 Approximations for Variance and Standard Deviation

As for the approximations in this case, we can estimate the standard deviation through calculating variance in two ways; one of them is direct,  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ , and the other one is the first-order Taylor expansion,  $\text{Var}(R/S) \approx \frac{\mu_R^2}{\mu_S^2} \left[ \frac{\sigma_R^2}{\mu_R^2} - 2 \frac{\text{Cov}(R,S)}{\mu_R \mu_S} + \frac{\sigma_S^2}{\mu_S^2} \right]$ , which is an approximation as opposed to the former.

First, we will try using the first formula, which would be in our case:

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

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 approximation for the ratio of two random variables. So far, we tried to use the former one. This approximation for the standard deviation of the center of mass in this case turned out to be

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}}. \quad (4.4.1)$$

In other words, this estimation comes from a first-order approximation of the expected value of the center of mass squared in the formula for variance,  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ . We will now prove the above equation. For our case, we have

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

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

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

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

where

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

and

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

as calculated in the previous chapter. Thus, we have

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

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

$$\begin{aligned} \text{Var}(CM) &\approx \frac{1}{4} \left( \frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2} \right) - \frac{1}{4} = \frac{1}{4} \left( \frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2} - 1 \right) = \\ &= \frac{1}{4} \left( \frac{16\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2} - \frac{12\alpha_2 + 12n\alpha_1^2}{12\alpha_2 + 12n\alpha_1^2} \right) = \frac{1}{4} \cdot \frac{4\alpha_2}{12(\alpha_2 + n\alpha_1^2)} = \frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}. \end{aligned}$$

which approaches 0 as  $n$  gets larger. Also,

$$\text{SD}(CM) = \sqrt{\text{Var}(CM)} \approx \sqrt{\frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}}.$$



This result will approach 0 as  $n$  goes to infinity. We will call this approximation  $s_1$  to make it easier to refer to it when we analyze our results.

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}}. \quad (s_1)$$

This approximation goes to 0 when  $n$  approaches infinity:

$$\lim_{n \rightarrow \infty} s_1 = \lim_{n \rightarrow \infty} \sqrt{\frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}} = 0.$$

Perhaps a thing to consider in the future is to estimate the expected value by using second-order Taylor expansion to see whether that approximation would prove to be more accurate than the others. However, as we saw in the case of finding the standard deviation of the center of mass in the one-dimensional discrete configuration, calculating the covariance term in the Taylor approximation formula was very complicated.

As for the second way to estimate variance, let us look at the first-order Taylor approximation for the ratio of two random variables,

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

where  $R = \sum_{i=0}^n Y_i M_i$  and  $S = \sum_{i=0}^n M_i$ . It turns out that the resulting approximation is the following:

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(n+1)\alpha_1^2}}. \quad (4.4.2)$$

We will call this approximation  $s_2$ . Let us go through the steps that led us to this result. Since we calculated needed components of the first-order Taylor approximation formula in the expected value approximation section, we have

$$\mu_R = \frac{1}{2}(n+1)\alpha_1,$$

$$\mu_S = (n+1)\alpha_1,$$

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

$$\sigma_S^2 = (n+1)(\alpha_2 - \alpha_1^2),$$

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

Therefore,

$$\begin{aligned} \text{Var}(R/S) &\approx \frac{\frac{1}{4}(n+1)^2\alpha_1^2}{(n+1)^2\alpha_1^2} \left[ \frac{(n+1)(\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2)}{\frac{1}{4}(n+1)^2\alpha_1^2} - 2\frac{\frac{1}{2}(n+1)[\alpha_2 - \alpha_1^2]}{\frac{1}{2}(n+1)\alpha_1(n+1)\alpha_1} + \frac{(n+1)(\alpha_2 - \alpha_1^2)}{(n+1)^2\alpha_1^2} \right] = \\ &= \frac{1}{4} \left[ \frac{\frac{1}{3}(\alpha_2 - \frac{1}{4}\alpha_1^2)}{\frac{1}{4}(n+1)\alpha_1^2} - 2\frac{\alpha_2 - \alpha_1^2}{(n+1)\alpha_1^2} + \frac{\alpha_2 - \alpha_1^2}{(n+1)\alpha_1^2} \right] = \frac{\frac{1}{3}\alpha_2 - \frac{1}{4}\alpha_1^2}{(n+1)\alpha_1^2} - \frac{\frac{1}{2}(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} + \frac{\frac{1}{4}(\alpha_2 - \alpha_1^2)}{(n+1)\alpha_1^2} = \\ &= \frac{\frac{4}{12}\alpha_2 - \frac{3}{12}\alpha_1^2 - \frac{6}{12}\alpha_2 + \frac{6}{12}\alpha_1^2 + \frac{3}{12}\alpha_2 - \frac{3}{12}\alpha_1^2}{(n+1)\alpha_1^2} = \frac{\frac{1}{12}\alpha_2}{(n+1)\alpha_1^2} = \frac{\alpha_2}{12(n+1)\alpha_1^2}. \end{aligned}$$

The standard deviation would thus be

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(n+1)\alpha_1^2}}. \quad (s_2)$$

Similarly to the first approximation, the estimated standard deviation from this formula goes to 0 as  $n$  approaches infinity, i.e.,

$$\lim_{n \rightarrow \infty} s_2 = \lim_{n \rightarrow \infty} \sqrt{\frac{\alpha_2}{12(n+1)\alpha_1^2}} = 0.$$

## 4.5 Simulations and Exact Value

In this section we will take a look at the results obtained from simulations in R, as well as those we get from using the exact value formulas.

We have derived a theoretical proof that the expected value of the center of mass in the one-dimensional uniform case is  $\mathbb{E}[CM] = \frac{1}{2}$  (Theorem 4.2.1). Using R, we can create a histogram of the sampling distribution of the sample means. The code is provided in Appendix A.5. We set the number to repeat the sampling to be  $C = 10^3 - 1$  (which would be the number of our data sets). We let  $m = 1$ ,  $M = 10$ ,  $p = 0.5$  and  $n = 10^4$ . Below are two histograms that show the distribution of the sample means from each data set, with sample means divided into an appropriate number of bins on the  $x$ -axis and their frequency on the  $y$ -axis. The histograms clearly indicate that the center of the distribution is approximately  $\frac{1}{2}$ , which is more visible in the second graph.

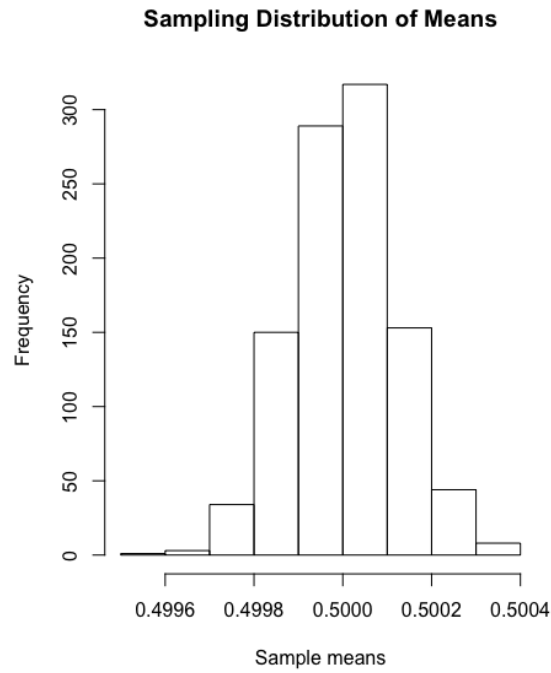


Figure 4.5.1. Sampling distribution of means of the expected value of the center of mass in the one-dimensional uniform case. Parameters:  $m = 1$ ,  $M = 10$ ,  $p = 0.5$ ,  $n = 10^4$ . Number of samplings:  $10^3 - 1$ .

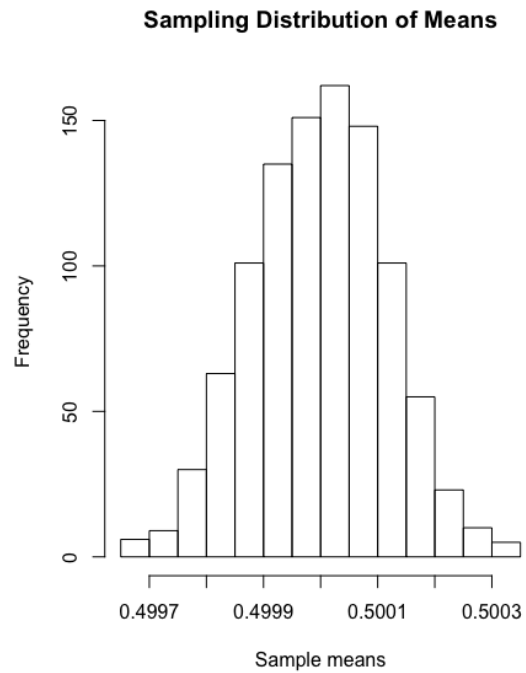


Figure 4.5.2. Sampling distribution of means of the expected value of the center of mass in the one-dimensional uniform case. Parameters:  $m = 1$ ,  $M = 10$ ,  $p = 0.5$ ,  $n = 10^4$ . Number of samplings:  $10^3 - 1$ .

As for the standard deviation of the center of mass in this case, we have results from repeated simulations ( $s$ ), one exact value formula that uses conditional expectation ( $\sigma$ ), and two approximations ( $s_1$  and  $s_2$ ). In this section, we will compare the results we can from simulations to those we get from the exact value formula to see how close they are. Below are the tables of both expected value and standard deviation obtained from simulations alone. These results are prone to variation.

<b>m = 1, M = 10</b>	<b>Expected Value from Simulations</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	0.499564	0.500186	0.499939
<b>0.2</b>	0.499467	0.500032	0.500044
<b>0.3</b>	0.499208	0.500412	0.500079
<b>0.4</b>	0.499587	0.500149	0.500212
<b>0.5</b>	0.499624	0.499910	0.500292
<b>0.6</b>	0.502131	0.500097	0.500034
<b>0.7</b>	0.499196	0.500130	0.500132
<b>0.8</b>	0.501184	0.499833	0.499848
<b>0.9</b>	0.497768	0.499552	0.499887

Figure 4.5.3. Expected value of the center of mass in the one-dimensional uniform case from simulations. Parameters used:  $m = 1$ ,  $M = 10$ ,  $p = 0.5$ ,  $n$  varied.

Expected value of the center of mass in the one-dimensional uniform case turns out to be very close to  $\frac{1}{2}$ , and as we saw, we have a theoretical proof that it is actually true that  $\mathbb{E}[CM] = \frac{1}{2}$ .

<b>m = 1, M = 10</b>	<b>Standard Deviation from Simulations</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	0.030169	0.009712	0.003011
<b>0.2</b>	0.030940	0.010253	0.003237
<b>0.3</b>	0.033391	0.010663	0.003353
<b>0.4</b>	0.035588	0.011450	0.003511
<b>0.5</b>	0.038262	0.011744	0.003734
<b>0.6</b>	0.040925	0.012578	0.003993
<b>0.7</b>	0.043203	0.013246	0.004302
<b>0.8</b>	0.047640	0.014763	0.004558
<b>0.9</b>	0.049542	0.015310	0.004853

Figure 4.5.4. Standard deviation of the center of mass in the one-dimensional uniform case from simulations. Parameters used:  $m = 1$ ,  $M = 10$ ,  $p = 0.5$ ,  $n$  varied.

For the same parameters, we will compute the standard deviation using the exact value formula. The results are as follows:

<b>m = 1, M = 10</b>	<b>Standard Deviation from Exact Value Formula</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	0.029972	0.009518	0.003011
<b>0.2</b>	0.031393	0.009965	0.003153
<b>0.3</b>	0.033029	0.010481	0.003315
<b>0.4</b>	0.034936	0.011082	0.003505
<b>0.5</b>	0.037188	0.011791	0.003730
<b>0.6</b>	0.039880	0.012642	0.003999
<b>0.7</b>	0.043103	0.013666	0.004323
<b>0.8</b>	0.046786	0.014862	0.004702
<b>0.9</b>	0.049446	0.015841	0.005015

Figure 4.5.5. Exact formula standard deviation of the center of mass in the one-dimensional uniform case. Parameters used:  $m = 1$ ,  $M = 10$ ,  $p = 0.5$ ,  $n$  varied.

The standard deviation obtained from exact value formula fits the data from the simulations very well. Similarly to the one-dimensional discrete case, we notice that if for the same values of  $p$ ,  $m$ , and  $M$  we increase the number of nodes  $n$  by  $x$ , then the standard deviation seems to decrease by  $\sqrt{x}$ . For instance, let us consider the data for  $m = 1$ ,  $M = 10$ ,  $p = 0.5$ . If we look at the ratio of the standard deviation for  $n = 1000$  to the one for  $n = 10000$ , which is an increase of the number of nodes  $n$  by a factor of 10, we have a decrease of

$$\frac{0.011791}{0.003730} \approx 3.16142037,$$

and

$$\sqrt{10} \approx 3.16227766.$$

The two numbers above are very close to each other, and it turns out to be the case if we were compare results for other parameters. We can find an explanation from this behavior from the approximations for the standard deviation:

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(\alpha_2 + n\alpha_1^2)}} \quad (s_1)$$

and

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(n+1)\alpha_1^2}}. \quad (s_2)$$

As opposed to the one-dimensional discrete case, the standard deviation decreases if we increase the number of nodes  $n$ . Again, we notice that  $\alpha_1 = mp + M(1-p)$  and  $\alpha_2 = m^2p + M^2(1-p)$  depend on the values of  $p$ ,  $m$  and  $M$  only, hence if we were to keep these constant (unchanged) and vary  $n$ , we can rewrite the second approximation as follows

$$\text{SD}(CM) \approx \sqrt{\frac{\alpha_2}{12(n+1)\alpha_1^2}} = \sqrt{\frac{1}{n+1} \cdot \frac{\alpha_2}{12\alpha_1^2}} = \sqrt{\frac{1}{n+1}} \cdot B,$$

where

$$B = \sqrt{\frac{\alpha_2}{12\alpha_1^2}}$$

is some constant. Therefore, if we increase  $n$  by  $x$ , then the new standard deviation would be

$$\text{SD}(CM) \approx \sqrt{\frac{1}{xn+1}} \cdot B,$$

meaning it decreases by

$$\frac{\sqrt{\frac{1}{n+1}} \cdot B}{\sqrt{\frac{1}{xn+1}} \cdot B} = \frac{\sqrt{\frac{1}{n+1}}}{\sqrt{\frac{1}{xn+1}}} = \sqrt{\frac{xn+1}{n+1}},$$

which is close to  $\sqrt{x}$  when  $n$  goes to infinity. For example, if we consider the standard deviation for  $n = 100$  and  $n = 1000$  (where  $n$  increases by 10), then the standard deviation decreases by

$$\sqrt{\frac{10 \cdot 100 + 1}{100 + 1}} \approx 3.14815677645,$$

which is very close to  $\sqrt{10} \approx 3.16227766$ . We can also conclude this from the first approximation, since the dominating term in the formula is the  $n$  in the numerator, thus indicating that if we increase  $n$  by some number, then the standard deviation decreases by the square root of the factor it increased by. As mentioned in the one-dimensional discrete case, these are only approximated changes because they are explained by approximations and not the exact value formula.

## 4.6 Analysis of Results

In this section, we will compare the standard deviation obtained from the exact value formula to the ones calculated from the approximations. We define the two approximations (4.4.1 and 4.4.2) in R and conduct simulations by varying  $M$  (one of the masses) to test which estimation suits the data from simulations (2nd column from the left in the tables below) and the exact value formula more (3rd column from the left). In order to avoid variation of data, we took a mean of five to ten simulation results. The R code is provided in Appendix A.2.

We compare the estimations for the standard deviation by calculating relative errors with respect to the numbers we get from the exact value formula ( $\sigma$ ). The 6th column from the left compares the approximation  $s_1$  to  $\sigma$ , and the 7th column compares the approximation  $s_2$  to  $\sigma$ . The last column (on the very right) depicts the ratio  $|\sigma - s_1|/|\sigma - s_2|$ , i.e., it compares which approximation gives results closer to the actual value ( $\sigma$ ). If the ratio is greater than 1, then the second approximation ( $s_2$ ) is closer to the standard deviation obtained through the exact value formula ( $\sigma$ ), and if the ratio is less than 1, then the first approximation ( $s_1$ ) yields values closer to the actual value. Below are two of the tables summarizing the data with results for  $m = 1$ , varied  $M$ 's,  $p = 0.5$ , and  $n = 100$ ,  $n = 1000$ ,  $n = 10000$ .

M	Standard Deviation				Relative Error in Approx.		Comparison
	Simulation (s)	Exact Value ( $\sigma$ )	Approx. 4.4.1 (s1)	Approx. 4.4.2 (s2)	$ \sigma - s_1  / s_1$	$ \sigma - s_2  / s_2$	
0.001	0.0412359	0.0407866	0.0403830	0.0405817	0.0099943	0.0050491	1.969741337
0.01	0.0393882	0.0404105	0.0400321	0.0402220	0.0094524	0.0046865	2.007427056
0.1	0.0372227	0.0371882	0.0369911	0.0371135	0.0053283	0.0020127	2.638554217
1	0.0282322	0.0287242	0.0287242	0.0287242	0.0000000	0.0000000	-
10	0.0371405	0.0371882	0.0369911	0.0371135	0.0053283	0.0020127	2.638554217
100	0.0404967	0.0404105	0.0400321	0.0402220	0.0094524	0.0046865	2.007427056
1000	0.0411688	0.0407866	0.0403830	0.0405817	0.0099943	0.0050491	1.969741337

Figure 4.6.1. Standard deviation of the center of mass in the one-dimensional uniform case, where  $m = 1$  is one of the masses,  $M$  is the other mass,  $p = 0.5$  is the probability, and  $n = 100$  is the number of nodes. The standard deviation from the simulations is denoted by ( $s$ ), the exact value is denoted by  $\sigma$ , the approximation 4.4.1 is denoted by  $s_1$ , and the approximation 4.4.2, is denoted by  $s_2$ .

Below are the same tables as the one above, but for  $n = 1000$  and  $n = 10000$ . For the latter, we have to use Mathematica to compute the standard deviation from the exact value formula as it contains a summation for large  $n$ , which cannot be computed in RStudio.

M	Standard Deviation				Relative Error in Approx.		Comparison
	Simulation (s)	Exact Value ( $\sigma$ )	Approx. 4.4.1 (s1)	Approx. 4.4.2 (s2)	$ \sigma - s1  / s1$	$ \sigma - s2  / s2$	
0.001	0.0129305	0.0128970	0.0128842	0.0128906	0.0009935	0.0004965	2
0.01	0.0129218	0.0127823	0.0127702	0.0127764	0.0009475	0.0004618	2.050847458
0.1	0.0118027	0.0117913	0.0117850	0.0117890	0.0005346	0.0001951	2.739130435
1	0.0091072	0.0091242	0.0091242	0.0091242	0.0000000	0.0000000	-
10	0.0118019	0.0117913	0.0117850	0.0117890	0.0005346	0.0001951	2.739130435
100	0.0126845	0.0127823	0.0127702	0.0127764	0.0009475	0.0004618	2.050847458
1000	0.0129716	0.0128970	0.0128842	0.0128906	0.0009935	0.0004965	2

Figure 4.6.2. Standard deviation of the center of mass in the one-dimensional uniform case, where  $m = 1$  is one of the masses,  $M$  is the other mass,  $p = 0.5$  is the probability, and  $n = 1000$  is the number of nodes.

M	Standard Deviation				Relative Error in Approx.		Comparison
	Simulation (s)	Exact Value ( $\sigma$ )	Approx. 4.4.1 (s1)	Approx. 4.4.2 (s2)	$ \sigma - s1  / s1$	$ \sigma - s2  / s2$	
0.001	0.0040417	0.0040784	0.0040780	0.0040782	0.0000981	0.0000490	2
0.01	0.0039671	0.0040423	0.0040419	0.0040421	0.0000940	0.0000470	2
0.1	0.0037398	0.0037298	0.0037296	0.0037297	0.0000536	0.0000214	2.5
1	0.0028661	0.0028866	0.0028866	0.0028866	0.0000010	0.0000010	1
10	0.0037259	0.0037297	0.0037296	0.0037297	0.0000530	0.0000208	2.546624687
100	0.0039716	0.0040422	0.0040419	0.0040421	0.0000937	0.0000467	2.006770224
1000	0.0040395	0.0040784	0.0040780	0.0040782	0.0000993	0.0000503	1.975833429

Figure 4.6.3. Standard deviation of the center of mass in the one-dimensional uniform case, where  $m = 1$  is one of the masses,  $M$  is the other mass,  $p = 0.5$  is the probability, and  $n = 10000$  is the number of nodes.

It seems that the approximations yield very similar results to the one we obtain by using the formula we proved. The estimate that uses first-order Taylor expansion for variance ( $s_2$ ) proved to be more accurate than the estimate we get from using second-order Taylor expansion on the expected value in the  $\text{Var}(CM) = \mathbb{E}[CM^2] - \mathbb{E}[CM]^2$  formula ( $s_1$ ). What we notice, again, is that as  $n$  goes to infinity, the standard deviation goes to 0.



Although our theoretical prove of the standard deviation formula proved to be accurate with outcomes we got from simulations, it seems that it is rather cumbersome and time-consuming to calculate it because of the summation term, especially with a large number of nodes. As for the approximations, they suggest that in this uniform case,  $\mathbb{E}[CM] \approx \frac{1}{2}$  and  $SD(CM) \rightarrow 0$  as  $n$  gets larger and larger, but a solid proof would be needed to confirm this. Thus, it would have been helpful if the approximations had an error term.

Another idea we can consider is comparing the standard deviation of the center of mass of the discrete, but standardized case, and the uniform case from this chapter. To standardize the discrete case, we use the formula

$$CM = \frac{\sum_{i=0}^n (i/n)M_i}{\sum_{i=0}^n M_i}.$$

Thus, we can calculate the standard deviations of the two cases using the exact value formula, and are 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:

<b>P = 0.5, m = 1, M = 10</b>	<b>Standard Deviation</b>		
<b>n</b>	<b>Discrete Case (d)</b>	<b>Uniform Case (u)</b>	<b>Difference (d-u)</b>
10	0.0854351	0.114479	-0.0290439
50	0.0344028	0.0524376	-0.0180348
100	0.0239734	0.0371882	-0.0132148
250	0.01502768	0.02356170	-0.008534020
500	0.01059444	0.01667050	-0.006076060
750	0.00864168	0.0136141	-0.004972420
1000	0.00748018	0.0117913	-0.00431112

Figure 4.6.4. Exact value standard deviation of the center of mass in the standardized discrete case (d) and the uniform case (u), where probability  $p = 0.5$ , masses are  $m = 1$  and  $M = 10$ , and number of nodes  $n$  is varied. The fourth column represents the difference between the standard deviation of the discrete and uniform cases.

As we can see, the standard deviation of the two cases is very similar. All of the times, the standard deviation of the uniform case is greater than that of the discrete case. However, the difference seems to decrease as  $n$  increases.

# 5

## Center of Mass in the Two-dimensional Uniform Case

### 5.1 Introduction to the Case

The last but not least interesting case in this paper is the two-dimensional uniform case. We consider at a unit circle on an  $xy$ -plane with polar coordinates, where the angle  $\theta$  is determined uniformly. The coordinates  $(x, y)$  are calculated by setting  $x = \cos \theta$  and  $y = \sin \theta$ . We define the center of mass as follows

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

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

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

$$SD(CM) \approx (0, 0).$$

We will try to prove these in the next sections.

In this case, analogically to the previous cases, we have that

$$\mathbb{E}[M_\theta] = \alpha_1,$$

$$\mathbb{E}[M_\theta^2] = \alpha_2.$$

## 5.2 Approximation for Expected Value

We can approximate the expected value by first-order Taylor approximation,  $\mathbb{E}[R/S] \approx \frac{\mathbb{E}[R]}{\mathbb{E}[S]}$ . It turns out that

$$\mathbb{E}[CM] \approx (0, 0). \quad (5.2.1)$$

To prove this, we first calculate needed components:

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

Thus, for the x-coordinate, we have:

$$\begin{aligned} \mathbb{E}[CM_X] &\approx \frac{\mathbb{E}[\sum_{i=0}^{2\pi} \cos(\theta) M_\theta]}{\mathbb{E}[\sum_{i=0}^{2\pi} M_\theta]} = \frac{\sum_{i=0}^{2\pi} \mathbb{E}[\cos(\theta)] \mathbb{E}[M_\theta]}{\sum_{i=0}^{2\pi} \mathbb{E}[M_\theta]} = \frac{\sum_{i=0}^{2\pi} \mathbb{E}[\cos(\theta)] \alpha_1}{n \alpha_1} = 0, \\ \mathbb{E}[CM_Y] &\approx \frac{\mathbb{E}[\sum_{i=0}^{2\pi} \sin(\theta) M_\theta]}{\mathbb{E}[\sum_{i=0}^{2\pi} M_\theta]} = \frac{\sum_{i=0}^{2\pi} \mathbb{E}[\sin(\theta)] \mathbb{E}[M_\theta]}{\sum_{i=0}^{2\pi} \mathbb{E}[M_\theta]} = \frac{\sum_{i=0}^{2\pi} \mathbb{E}[\sin(\theta)] \alpha_1}{n \alpha_1} = 0. \end{aligned}$$

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

## 5.3 Approximation for Variance and Standard Deviation

In order to find variance of the center of mass of this configuration, we will use the formula  $\text{Var}(CM) = \mathbb{E}[CM^2] - \mathbb{E}[CM]^2$ . We can approximate the former term for the  $x$ -coordinate as

$$\mathbb{E}[CM_X^2] = \mathbb{E} \left[ \frac{(\sum \cos(\theta) M_\theta)^2}{(\sum M_\theta)^2} \right] \approx \frac{\mathbb{E} \left[ (\sum \cos(\theta) M_\theta)^2 \right]}{\mathbb{E} \left[ (\sum M_\theta)^2 \right]}.$$

The  $y$ -coordinate would have  $\sin(\theta)$  instead of  $\cos(\theta)$ . The approximated standard deviation of the center of mass in this two-dimensional uniform case turns out to be

$$\text{SD}(CM) \approx (0, 0), \quad (5.3.1)$$

regardless of all variables:  $n$ ,  $m$ ,  $M$ ,  $p$ . Let us prove this claim. We know that

$$\mathbb{E}[\cos^2(\theta)] = \frac{1}{2\pi} \int_0^{2\pi} \cos^2(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 + \cos(2\theta)}{2} d\theta = \frac{1}{2\pi} \left[ \frac{1}{2}\theta \Big|_0^{2\pi} + \frac{\sin(2\theta)}{4} \Big|_0^{2\pi} \right] = \frac{1}{2\pi} [\pi + 0 + 0] = \frac{1}{2}.$$

Similarly,

$$\mathbb{E}[\sin^2(\theta)] = \frac{1}{2}.$$

Thus, for each of the coordinates, we have

$$\begin{aligned} \mathbb{E}[CM_X^2] &\approx \frac{\mathbb{E} \left[ (\sum \cos(\theta) M_\theta)^2 \right]}{\mathbb{E} \left[ (\sum M_\theta)^2 \right]} = \frac{\mathbb{E} \left[ \sum \cos^2(\theta) M_\theta^2 + \sum_{\theta \neq \delta} \cos(\theta) \cos(\delta) M_\theta M_\delta \right]}{\mathbb{E} \left[ \sum M_\theta^2 + \sum_{\theta \neq \delta} M_\theta M_\delta \right]} = \\ &= \frac{\mathbb{E} \left[ \frac{1}{2} \alpha_2 \right]}{\mathbb{E} \left[ (n+1) \alpha_2 + n(n+1) \alpha_1^2 \right]} = \frac{\alpha_2}{2(n+1)(\alpha_2 + n\alpha_1^2)}. \end{aligned}$$

Clearly, the expected value of the center of mass of the  $x$ -coordinate goes to 0 as  $n$  goes to infinity. As for the  $y$ -coordinate, we have

$$\begin{aligned} \mathbb{E}[CM_Y^2] &\approx \frac{\mathbb{E} \left[ (\sum \sin(\theta) M_\theta)^2 \right]}{\mathbb{E} \left[ (\sum M_\theta)^2 \right]} = \frac{\mathbb{E} \left[ \sum \sin^2(\theta) M_\theta^2 + \sum_{\theta \neq \delta} \sin(\theta) \sin(\delta) M_\theta M_\delta \right]}{\mathbb{E} \left[ \sum M_\theta^2 + \sum_{\theta \neq \delta} M_\theta M_\delta \right]} = \\ &= \frac{\mathbb{E} \left[ \frac{1}{2} \alpha_2 \right]}{\mathbb{E} \left[ (n+1) \alpha_2 + n(n+1) \alpha_1^2 \right]} = \frac{\alpha_2}{2(n+1)(\alpha_2 + n\alpha_1^2)}. \end{aligned}$$

Thus, as  $n$  goes to infinity,  $\mathbb{E}[CM_X^2]$ ,  $\mathbb{E}[CM_Y^2]$  both go to 0. As a result, variance, as well as standard deviation of the center of mass, will approach 0 since  $\mathbb{E}[CM_X] = \mathbb{E}[CM_Y] \approx 0$ , that is,

$$\text{Var}(CM) \approx (0, 0).$$

As a result,

$$\text{SD}(CM) \approx (0, 0).$$

## 5.4 Simulations

Below are tables with expected value and standard deviation of the center of mass in the two-dimensional uniform case. We set the values of small mass  $m = 1$ , big mass  $M = 10$ , and then we vary the number of nodes  $n$  as well as the probability  $p$ .

<b>m = 1, M = 10</b>	<b>Expected Value from Simulations</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	(0.0222017, 0.0666151)	(-0.00936532, -0.00730835)	(0.00604702, 0.00106627)
<b>0.2</b>	(-0.10423, 0.0501832)	(0.00338956, -0.00796426)	(0.00576922, -0.0027828)
<b>0.3</b>	(-0.107908, 0.0192416)	(-0.00996234, -0.00722221)	(-0.00298045, -0.00140758)
<b>0.4</b>	(0.0040302, 0.0794788)	(-0.0511133, -0.0163147)	(0.0121272, -0.00523938)
<b>0.5</b>	(0.0077282, -0.0635334)	(-0.0556507, -0.0298255)	(0.0210253, -0.0374929)
<b>0.6</b>	(-0.0563869, -0.00132405)	(0.00918773, -0.0168086)	(-0.00196928, -0.0135746)
<b>0.7</b>	(-0.128005, 0.000296738)	(-0.0133626, -0.00531741)	(0.00542514, 0.015403)
<b>0.8</b>	(0.0239391, 0.0341421)	(-0.0158461, -0.00423165)	(-0.00953162, -0.002967)
<b>0.9</b>	(-0.0846125, 0.040083)	(0.123384, -0.0174753)	(-0.00518863, 0.00184773)

Figure 5.4.1. Expected value for the two-dimensional uniform case from simulations, where  $m = 1$  is one of the masses,  $M = 10$  is the other mass,  $p$  is the probability, and  $n$  is the number of nodes.

<b>m = 1, M = 10</b>	<b>Standard Deviation from Simulations</b>		
<b>P</b>	<b>n = 100</b>	<b>n = 1000</b>	<b>n = 10000</b>
<b>0.1</b>	(0.0208283, 0.021557)	(0.00643735, 0.00671376)	(0.0021101, 0.00210842)
<b>0.2</b>	(0.0324626, 0.0317474)	(0.0100051, 0.00995723)	(0.00309201, 0.00319267)
<b>0.3</b>	(0.0401742, 0.0393795)	(0.0125441, 0.0125108)	(0.00406813, 0.0040855)
<b>0.4</b>	(0.0475032, 0.0479244)	(0.01526, 0.0151797)	(0.00486527, 0.00478135)
<b>0.5</b>	(0.0610687, 0.0560359)	(0.0179913, 0.0190948)	(0.00577807, 0.00581872)
<b>0.6</b>	(0.0696184, 0.0653889)	(0.0226444, 0.0212252)	(0.0066838, 0.00703437)
<b>0.7</b>	(0.0799278, 0.0775909)	(0.0250466, 0.024663)	(0.00771198, 0.0082563)
<b>0.8</b>	(0.0967828, 0.0870833)	(0.0285229, 0.0295491)	(0.00866917, 0.00910544)
<b>0.9</b>	(0.103957, 0.0980932)	(0.100674, 0.0988582)	(0.0100682, 0.0101945)

Figure 5.4.2. Standard deviation for the two-dimensional uniform case from simulations, where  $m = 1$  is one of the masses,  $M = 10$  is the other mass,  $p$  is the probability, and  $n$  is the number of nodes.

Although the recorded results are prone to variation, we can still conclude that the expected value of the center of mass is approximately  $(0, 0)$ , and so is the standard deviation. It also seems to be the case that as  $n$  increases, the results get closer to the origin of the  $xy$ -plane - the point  $(0, 0)$ . Hence, the simulations confirm our approximations that  $\mathbb{E}[CM] \approx (0, 0)$  and  $SD(CM) \approx (0, 0)$ .

It would be helpful to find an exact value formula for this two-dimensional uniform case so that the results could be theoretically proven.

# 6

## Future Research

Based on the content of this project, we can state a couple of conjectures for future research.

One thing to consider is simplifying the exact value formulas in the one-dimensional discrete and uniform cases so that they are easier to compute. Perhaps this could help us with understanding the behavior of the center of mass, its expected value as well as standard deviation.

From what we have seen, we can state the following conjecture:

**Conjecture 1.** *The standard deviation of the center of mass in the one-dimensional uniform case approaches 0 as  $n$  goes to infinity, regardless of other variables  $m$ ,  $M$  and  $p$ .*

It would be nice to have a formula that does not involve summations, or binomials. Since Mathematica simplifies the exact value formulas into those that contain generalized hypergeometric functions, a further study into what kind of functions these are and what characteristics they have might give some insight into whether the limit of the standard deviation of the center of mass goes to 0 in the one-dimensional uniform case.

Another aspect of this project, related to the previous conjecture, that could be part of future research is distribution of the center of mass in the one-dimensional uniform case as  $n$  goes to infinity. The expected value of the center of mass is  $\frac{1}{2}$ , and it seems like the standard deviation goes to 0 as  $n$  approaches infinity, which implies that the distribution converges to a point mass distribution. In the case considered, for each  $n$  we have different probability distributions of the

center of mass that can be expressed as functions. If we fix the values of  $m$ ,  $M$  and  $p$ , we expect the probability and the distribution functions of the center of mass as  $n$  goes to infinity to be as follows:

$$f(x) = \lim_{n \rightarrow \infty} P(CM = x) = \begin{cases} 1 & \text{if } x = \frac{1}{2}, \\ 0 & \text{otherwise} \end{cases}$$

and

$$F(x) = \lim_{n \rightarrow \infty} P(CM \leq x) = \begin{cases} 0 & \text{if } x < \frac{1}{2}, \\ 1 & \text{if } x \geq \frac{1}{2}. \end{cases}$$

We can claim the following conjecture:

**Conjecture 2.** *The center of mass in the one-dimensional uniform case converges to a point-mass distribution with mean  $\frac{1}{2}$  and standard deviation of 0 as  $n$  goes to infinity.*

Perhaps it is possible to prove in a similar way that the Central Limit Theorem is proved (for instance by using moment generating functions) that all these probability functions of the center of mass converge into a point-mass distribution with mean  $\frac{1}{2}$  and standard deviation 0.

Last but not least, there is a plethora of other configurations that are interesting, and whose expected value and standard deviation of the center of mass could prove to be significant in the analysis of the center of mass of complex, or high-dimensional systems. An example could be a two-dimensional configuration with a unit square with vertices  $(0, 0)$ ,  $(1, 0)$ ,  $(1, 1)$  and  $(0, 1)$  on the  $xy$ -coordinate plane. The masses could be assigned uniformly inside the square, or on the edges of the square. It is also possible to assign the masses in a different way — for instance, instead of assigning the positions from a uniform distribution, perhaps a normal distribution could be used. There are also other known distributions with interesting properties, such as the exponential distribution, gamma distribution or chi-squared distribution. Furthermore, more masses could be added to the system, say,  $m_1, m_2, m_3$ . There are plenty variations of the configurations, as well as ways of adding other variables which would make the cases even more intriguing in future research.

# Appendix A

## R Codes

### A.1 One-dimensional Discrete Case

```
# Center of Mass: One-dimensional Discrete Case

# Setup

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

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

# Probability of getting m
p <- 0.5

# Number of nodes
N <- 10^3

# Create a vector with positions
positions <- 0:N

# Simulation
# Use the sample() function to sample from a vector of masses m and M.

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

# Creating data sets
for(i in 1:B)
{
```



```

myindex <- sample(samplespace, size = N+1, replace = TRUE, prob = c(p
,1-p))
result[i] <- sum(((positions)*index))/sum(index)
}

# Calculating the exact value formula
k <- c(0:N+1)
va.totM <- ((N-1)*(3*N+2)/12) - ((N^2)/4) + sum(((choose(N+1,k)*(p^k)*
(1-p)^(N+1-k))/(((k*m+(N+1-k)*M))^2))*((N+1)*(N+2)/12)*(k*(m^2)+(N+1-
k)*(M^2)))
std.totM <- sqrt(va.totM)
# Standard deviation of the center of mass from simulations
s <- sd(result)

# Calculating approximations for standard deviation
sdd <- function(M,m,p,N){
  L <- function(pow,i,M,m,p){
    alpha <- (((m^i)*p)+(M^i)*(1-p))^pow
    return(alpha)
  }
  # Approximation a1
  std.tay <- sqrt(((1/12)*(N*(N+2)*(L(1,2,M,m,p)-L(2,1,M,m,p)))/(L(1,2,M,
m,p)+N*L(2,1,M,m,p))))
  # Approximation a2
  std.form <- sqrt(((1/12)*((N*(N+2)*(L(1,2,M,m,p)-L(2,1,M,m,p)))/((N+1)*
L(2,1,M,m,p))))
  paste("Taylor/Approx: SD(CM)=", signif(std.tay), "Formula/Approx.: SD(C
M)=", signif(std.form))
}

```

## A.2 One-dimensional Uniform Case

```

# Center of Mass - One-dimensional Uniform Case

# Setup

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

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

# Probability of getting m
p <- 0.5

# Number of nodes
N <- 10^3

# Simulation
# Use the sample() function to sample from a vector of masses m and M.

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

# Creating data sets
for(i in 1:B)
{
  random.numbers <- runif(N+1, 0, 1) # Vector with positions, uniform
  from 0 to 1
  myindex <- sample(samplespace, size = N+1, replace = TRUE, prob = c(p
    ,1-p))
  result[i] <- sum(((random.numbers)*myindex))/sum(myindex)
}

# Calculating the exact value formula
k <- c(0:N+1)
va.totM <- sum((((choose(N+1,k)*(p^k)*(1-p)^(N+1-k))/(((k*m+(N+1-k)*M)
  ^2))*(1/12)*(k*(m^2)+(N+1-k)*(M^2))))
std.totM <- sqrt(var.totM)

# Calculating approximations for expected value and standard deviation
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 approximated expected value using second-order Taylor
  expansion:
ex.taylor <- (mur/mus) - (covrs/mus^2) + (mur*vars/mus^3)
# Standard deviation approximation s2:
va.taylor <- (mur^2/mus^2)*((varr/mur^2) - 2*(covrs/(mur*mus)) + (vars
  /mus^2))
std.taylor <- sqrt(va.taylor)
# Standard deviation approximation s1:
va.form <- (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)
std.form <- sqrt(va.form)
paste("Taylor: E[CM]=", signif(ex.taylor), "SD(CM)=", signif(std.
  taylor), "Formula/Approx.: SD(CM)=", signif(std.form))
}
```

## A.3 Standard Deviation: Discrete vs Uniform Case

```

# Center of Mass: One-dimensional Standardized Discrete vs Uniform Case

# Setup

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

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

# Probability of getting m
p <- 0.5

# Number of nodes
N <- 10^3

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

# Simulation
# Use the sample() function to sample from a vector of masses m and M.

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

# Creating data sets
for(i in 1:B)
{
  random.numbers <- runif(N, 0, 1) # Vector with positions for the
  continuous case, uniform from 0 to 1.
  index <- sample(samplespace, size = N, replace = TRUE, prob = c(p,1-p)
  )
  resultd[i] <- sum(((positions/N)*index))/sum(index) # Results for the
  standardized discrete case
  resultc[i] <- sum(((random.numbers)*index))/sum(index) # Results for
  the uniform case
}

paste("Discrete: E[CM]=", signif(mean(resultd)), "Continuous: E[CM]=",
  signif(mean(resultc)))
paste("Discrete: SD(CM)=", signif(sd(resultd)), "Continuous: SD(CM)=",
  signif(sd(resultc)))

```

## A.4 Two-dimensional Uniform Case

```
# Center of Mass: Two-dimensional Uniform case (unit circle)

# Setup

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

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

# Probability of getting m
p <- 0.5

# Number of nodes
N <- 10^3

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

# Simulation
# Use the sample() function to sample from a vector of masses m and M.

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

# Creating data sets
for(i in 1:B)
{
  theta <- runif(N, 0, 2*pi) # Create a vector with positions
  index <- sample(samplespace, size = N, replace = TRUE, prob = c(p,1-p)
  )
  result1[i] <- sum((x*index))/sum(index)
  result2[i] <- sum((y*index))/sum(index)
}

# Expected value:
m1 <- mean(result1) # x-coordinate
m2 <- mean(result2) # y-coordinate

# Standard Deviation:
s1 <- sd(result1) # x-coordinate
s2 <- sd(result2) # y-coordinate
```

## A.5 Sampling Distribution of Sample Means

```

# Center of Mass - One-dimensional Uniform Case

# Sampling distribution of the sample means

# Setup

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

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

# Probability of getting m
p <- 0.5

# Number of nodes
N <- 10^4

# Simulation
# Use the sample() function to sample from a vector of masses m and M.

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

# Creating data sets
for(i in 1:B) {
  for(j in 1:B)
  {
    random.numbers <- runif(N+1, 0, 1) # Vector with positions, uniform
      from 0 to 1
    myindex <- sample(samplespace, size = N+1, replace = TRUE, prob = c(
      p,1-p))
    result[j] <- sum(((random.numbers)*myindex))/sum(myindex)
  }
  samplingmean[i] <- mean(result)
}

hist(samplingmean, xlab = "Sample means", main = "Sampling Distribution
of Means")

```



# Appendix B

## Mathematica Codes

### B.1 Algorithm for Calculating Double, Triple and Quadruple Summations

In order to calculate the double, triple and quadruple summations, we will use the **Sum**[*f*, *i*, *imin*, *imax*, *j*, *jmin*, *jmax*, ...] command that evaluates the multiple sum. Since we are mostly struggling to find the summation in cases where the variables do not equal each other (for instance,  $i \neq j \neq k \neq l$ ), we also have to use the **Boole**[*expr*] command, which yields 1 if *expr* is True and 0 if it is False., inside the **Sum** function to make sure all the conditions are met.

First, we have to write the term we want to sum up in terms of a function whose input is the variables of interest, and then we include it into the **Sum** command.

Below is a part of the Mathematica code used with chosen functions.

```
(** Double, Triple and Quadruple Summations **)
```

```
f[i-,j-]:=i * j
```

```
Sum[f[i,j]Boole[i ≠ j], {j, 0, n}, {i, 0, n}]
```

```
 $\frac{1}{12} (-2n - 3n^2 + 2n^3 + 3n^4)$ 
```

```
f[i-,j-]:=1
```



**Sum**[ $f[i, j]$ Boole[ $i \neq j$ ], { $j, 0, n$ }, { $i, 0, n$ }]

$$n + n^2$$

**f**[ $i, j, k$ ]:=1

**Sum**[ $f[i, j, k]$ Boole[ $i \neq j \neq k$ ], { $j, 0, n$ }, { $i, 0, n$ }, { $k, 0, n$ }]

$$\begin{cases} -\text{Ceiling}[n] + 3\text{Ceiling}[n]^2 - 2\text{Ceiling}[n]^3 - 3\text{Ceiling}[n]\text{Floor}[n] + 3\text{Ceiling}[n]^2\text{Floor}[n] & n > 0 \\ 0 & \text{True} \end{cases}$$

**f**[ $i, j, k, l$ ]:=1

**Sum**[ $f[i, j, k, l]$ Boole[ $i \neq j \neq k \neq l$ ], { $j, 0, n$ }, { $i, 0, n$ }, { $k, 0, n$ }, { $l, 0, n$ }]

$$\begin{cases} 2\text{Ceiling}[n] - 9\text{Ceiling}[n]^2 + 10\text{Ceiling}[n]^3 - 3\text{Ceiling}[n]^4 + 8\text{Ceiling}[n]\text{Floor}[n] \\ -12\text{Ceiling}[n]^2\text{Floor}[n] + 4\text{Ceiling}[n]^3\text{Floor}[n] & n > 0 \\ 0 & \text{True} \end{cases}$$

## B.2 Computing the Exact Value Formula for Large $n$

(\*\* SP - Exact Value Formula for sd uniform \*\*)

**f**[ $n, p, m, M$ ]:=

**Sqrt**[

$$\sum_{k=0}^{n+1} ((\text{Binomial}[n+1, k] * (p^k) * (1-p)^{(n+1-k)}) / ((k * m + (n+1-k) * M)^2) * (1/12) * (k * (m^2) + (n+1-k) * (M^2)))$$

(\*\* Computes the function for chosen values of  $n, p, m, M$  \*\*)

**f**[10<sup>4</sup>, 0.5, 1, 10]

(\*\* SP - Exact Value Formula for sd discrete \*\*)

**g**[ $n, p, m, M$ ]:=

**Sqrt**[(( $n-1$ ) \* ( $3n+2$ )/12) - ( $n^2/4$ )+

$$\sum_{k=0}^{n+1} ((\text{Binomial}[n+1, k] * (p^k) * (1-p)^{(n+1-k)}) / ((k * m + (n+1-k) * M)^2) * ((n+1)(n+2)/12) * (k * (m^2) + (n+1-k) * (M^2)))$$

(\*\* Computes the function for chosen values of  $n, p, m, M$  \*\*)

**g**[10<sup>4</sup>, 0.5, 1, 10]

# Bibliography

- [1] Laura M. Chihara, *Mathematical Statistics with Resampling and R*, John Wiley, 2011.
- [2] Geza Schay, *Introduction to Probability with Statistical Applications*, Birkhauser, Boston, 2007.
- [3] Howard Seltman, *Approximations for Mean and Variance of a Ratio*, available at <http://www.stat.cmu.edu/~hseltman/files/ratio.pdf>.
- [4] Alan Stuart and Keith Ord, *Kendall's Advanced Theory of Statistics*, Arnold, 1998.
- [5] Regina C. Elandt-Johnson, *Survival Models and Data Analysis*, John Wiley and Sons, 2014.
- [6] Gerhard Beutler, *Methods of Celestial Mechanics: Volume I: Physical, Mathematical, and Numerical Principles*, Springer Science and Business Media, 2005.
- [7] Finnegan Maximilan Muller Hardy, *The Expectation for the Center of Mass of Finite Integer Grids*, [http://digitalcommons.bard.edu/senproj\\_s2017/300](http://digitalcommons.bard.edu/senproj_s2017/300). Senior Projects Spring 2017.300.
- [8] Kazuhiko Aomoto and Michitake Kita, *Theory of Hypergeometric Functions*, Springer Science and Business Media, 2011.