Estimating the Probability That the Explosion of an Ink Sphere Produces a Dictionary

Braden Oh

March 7, 2023

Abstract

There is a thought experiment that compares the probability of life originating by random chance to the probability of a dictionary being printed in the explosion of a printing shop; this thought experiment has not been formally analyzed. This paper presents a highly simplified abstraction of the thought experiment and calculates the probability for the likelihood that an explosion of a sphere of ink results in a spray pattern that prints a 1989 Merriam-Webster English dictionary. The model presented consists of a sphere of ink exploding within a spherical shell composed of pieces of printing paper. The ink sphere is modeled as a composite of rectangular pyramid sectors that expand outward until their footprint perfectly overlaps a target piece of paper. Each sector is modeled as a stack of discrete layers of ink droplets, allowing the ink droplets within each sector to be quantified by a three-dimensional Riemann sum. The possible arrangements of these droplets are modeled by combinations and permutations that require the evaluation of enormous factorials to solve (e.g., 5,025,780!). These factorials are too large for a computer to evaluate by multiplication—or even Sterling's Approximation—due to limitations in compute time and decimal precision. As a result, an algorithm for using floating point numbers to conduct large combinatorial approximations in logarithmic space is derived and presented. The computed likelihood of a dictionary being printed within this model is $\approx 1.30 \times 10^{-817,692,555}$, or, in a more familiar form, ≈ 1 in $7.68 \times 10^{817,692,554}$.

1 Background and Assumed Model

There is a thought experiment that compares the probability of life originating by random chance to the probability of a dictionary being printed in the explosion of a printing shop. It is unclear who first posed this thought experiment. Academic analysis has been given to the probability of life as we know it arising [1][2][3]. Academic analysis has also been given to determine the probability of extremely low-likelihood events, such as the probability of a person quantum tunneling through a wall, the probability of infinite monkeys with typewriters producing the complete works of Shakespeare, or the probability of a fully intact brain appearing out of quantum fluctuations in the vacuum of space [4][5][6][7]. Academic literature studying the thought experiment of an explosion printing a dictionary, however, is not readily available. Although difficult to model the entire scenario of a print factory exploding, this paper explores a highly abstracted and idealized model with the purpose of providing an entry point for future academic exploration.

The model used in this paper consists of a sphere of ink exploding within a spherical shell comprised of pieces of printing paper; the distribution of ink particles is modeled by sectors of the sphere that expand outwards, eventually striking the walls of the paper shell. A diagram of this model is shown in Figure 1. In order to analyze this model mathematically, many simplifications must be made, including the following primary assumptions:

1. The ink sphere contains exactly the amount of ink necessary to print a single dictionary.
2. The roughly-spherical paper shell is comprised of the number of rectangular pieces of paper necessary to print a single dictionary.
3. The pieces of paper comprising the shell are at fixed locations in space.
4. All pages of the model dictionary contain the average number of pixels per page of a Merriam-Webster *Webster's Dictionary of English Usage.*
5. The dictionary is printed at a standard print resolution of 300 dots per inch (dpi).

6. At the instant of explosion, the ink sphere atomizes into identical droplets, each with the diameter to produce a 300 dpi pixel.
7. The sphere expands uniformly, resulting in an equal number of ink droplets per sector.
8. The sphere is analyzed as a set of identical, expanding sectors, with the number of sectors being equal to the number of pages in the dictionary and each containing enough ink to print one page.
9. Each expanded sector's footprint encloses a single piece of paper.
10. Each sector is a rectangular pyramid comprised of a stack of flat layers of ink droplets.
11. Ink droplets within any single layer are randomly arranged within the target footprint, but no two droplets in a single layer can land in the same location.
12. The order of the printed pages does not matter.

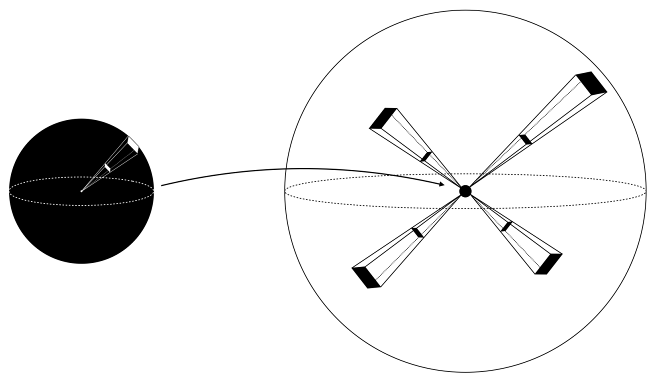

Figure 1: The model this paper uses is a sphere of ink comprised of sectors that are approximated by rectangular pyramids. When the explosion occurs, these sectors expand outward until they strike a roughly spherical shell comprised of pieces of paper. The footprint of each pyramid encloses a single piece of paper, ensuring the ink from that sector lands somewhere on that page.

The realistic nature of these assumptions is questionable but acceptable within reason for a highly approximated model. Although it is not possible to create a perfectly closed shell (polyhedron) with identical, rectangular faces, prior work in hemisphere partitioning has shown that it is possible to partition a hemisphere into equal areas with constrained aspect ratios that well approximate same-size rectangles [8]. Although the likelihood of an explosion atomizing ink droplets instantly and identically is low, the concept of flash boiling for ideal spray generation has been researched and has a potential application in the automotive engine industry for the purpose of optimizing fuel injection [9][10]. Although dot-matrix printers can produce resolutions as low as 60 dpi, such printers are no longer commonly used, so a resolution of 300 dpi (a typical medium to high quality resolution) is used by the model [11][12].

Although this model does not analyze an entire print factory, it does allow for quantitative analysis of the probability that a random explosion of ink particles within an idealized environment results in a Merriam-Webster English dictionary; it thus serves as an entry point for further academic analysis and discussion of this thought experiment.

2 Numbers of Pixels

The dictionary used in this model is the 1989 edition of the Merriam-Webster *Webster's Dictionary of English Usage*. To determine the average number of pixels per page (as required by assumption 4) 30 pages were selected at random from the 973 pages that make up the main body of this edition (ignoring brief preface and suffix sections, such as the edition notes and bibliography). These 30 pages were extracted from a portable document format (PDF) file of the dictionary, converted into black and white bitmaps, and had their pixels counted by a Python script. Of 149,261,354 total pixels counted, 19,476,801 were black, yielding an average of 13.0488% black pixels per page.

The dimensions of a page of the 1989 Merriam-Webster English dictionary are approximately 6.15×9.08 inches. At a resolution of 300^2 dots per square inch, each page is comprised of 5,025,780 pixels. Taking 13.05% of these pixels yields 655,803 black pixels (resulting in 638,096,319 total ink droplets necessary to print the entire dictionary). The average number of pixels per page will be denoted hereafter as $\Sigma N =$ 655,803 droplets.

3 Ink Droplet Physical Properties

By primary assumption 5 the dictionary is printed at a resolution of 300 dpi. By primary assumption 6 the ink sphere atomizes into identical droplets, each capable of yielding a pixel that measures $\frac{1}{300}$ of an inch, or $84.\overline{6}$ microns. Prior research in the microscopic topography of ink on paper has found that a typical layer of ink on a paper is approximately 2.5 microns deep [13]. Considering each pixel to be a cylinder with a diameter of $84.\overline{6}$ microns and a thickness of 2.5 micrometers, we can calculate the volume of ink in each pixel. This volume, in turn, equals the volume of ink in a spherical droplet of the ink spray, $V_d = 14\,075.21$ micron^3, which will have a diameter of $D_d = 29.96$ microns. The total volume of ink necessary to print a dictionary is therefore $V_d \times \Sigma N = 8.98$ cm^3. Before being exploded, this volume of ink occupies a sphere of radius $r_0 = 1.29$ cm.

4 Geometry of an Expanding Sector

A sector of the initial ink sphere is approximated by a rectangular pyramid that, by assumption 9, has an expanded footprint that covers one page of the dictionary (the entire initial ink sphere is comprised of 973 pyramids, one for each page of the dictionary). This is similar to the analysis of the footprint of a sensor with a rectangular field of vision, see [14]. By assumption 10, the rectangular pyramid is a composite of layers of ink particles, each with a distinct thickness of D_d. The pyramid's volume can thus be considered as the sum of the volumes of thin boxes, each of which has a rectangular base and a height of D_d; this is a volumetric Riemann sum similar to the one used to derive the formula for the volume of a pyramid for students of calculus [15][16]. A diagram of this geometry and relevant dimensions is shown in Figure 2.

The height of the expanded pyramid, or the distance from the center of the initial ink sphere to the center of the target piece of paper, is $h_0 = \sqrt{r_0^2 - (\frac{w_0}{2})^2}$, a function of the radius of the spherical paper shell, r_0, and the width of the target piece of paper, w_0. When taking a Riemann sum of the pyramid's volume along the h axis, the volume component at a particular step, n, is a rectangular prism with length l_n, width w_n, and thickness Δh (which is equal to the diameter of a single droplet of ink, D_d). The values of l_n and w_n can be found with similarity of triangles by multiplying l_0 and w_0 each by the ratio of the altitude of the pyramid at that step, h_n, to the expanded altitude, h_0. The Riemann sum approximating the total volume of the pyramid, ΣV, is the sum of all volume components:

$$\Sigma V = \sum_{n=0}^{m} \frac{w_0 \times l_0 \times \Delta h^3 \times n^2}{r_0^2 - (\frac{w_0}{2})^2} \tag{1}$$

The limit of the sum, m, is the integer number of layers in the pyramid rounded to the nearest integer: $m = \frac{h_0}{\Delta h}$. Assumption 10 models the ink in a sector as a series of layers comprised of ink droplets, and assumptions 1, 4, 7, and 8 establish that each sector contains exactly the number of droplets necessary to print a single page. The number of droplets in a particular layer, $N(n)$, is therefore given by

$$N(n) = \frac{V(n)}{\Sigma V} \Sigma N \tag{2}$$

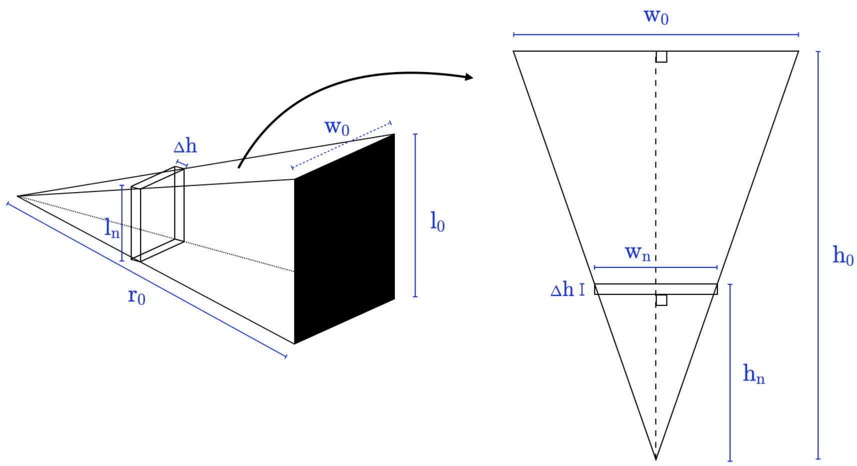

Figure 2: Each pyramidal sector is represented as a composition of slices, with each slice being the thickness of an ink droplet and containing a discrete number of ink particles. Left is a sector diagram with a discrete slice shown, and right is a top-view of the same geometry. The labeled values are used in the following set of calculations. w_0 and l_0 are the dimensions of a page, 6.15 and 9.08 inches (156,210 and 230,632 microns), respectively, and $\Delta h = D_d$.

where $V(n)$ is the volume of that layer of the sector,

$$V(n) = \frac{w_0 \times l_0 \times \Delta h^3 \times n^2}{r_0^2 - (\frac{w_0}{2})^2} \tag{3}$$

The parameters for evaluation of these expressions (applied to the initial, un-exploded ink sphere) are the following:

$$\begin{aligned} w_0 &= 15.62 \text{ cm} \\ l_0 &= 23.06 \text{ cm} \\ \Delta h &= 29.96 \text{ microns} \\ n &= 430 \text{ layers} \\ r_0 &= 1.29 \text{ cm} \end{aligned}$$

4.1 Computational Evaluation of $N(n)$, with Corrections

Equations 2 and 3 can be used to generate a set of values that denote the number of ink particles at each layer of the sector, $\{N(0), N(1), N(2), ...N(m)\}$. This set was calculated computationally in Python 3.8.5 and resulted in 655,795 total droplets across all layers. ΣN = 649,226 droplets, however, meaning the Riemann sum calculated an extraneous 6,569 droplets (an error of +1.01%). To correct for this error, each layer had subtracted from it a portion of these extraneous droplets proportional to the fraction of the total number of sector droplets contained within that layer. This resulted in the subtraction of 6,566 extraneous droplets and reduced the error in the total number of droplets to below +0.0005%.

5 Arrangements Within a Single Layer

By assumption 11, the ink droplets are randomly distributed throughout the target area; because all ink droplets are identical, the number of arrangements for droplets in a single layer can be modeled as a simple combination,

$$C(n) = \binom{L}{N(n)} \tag{4}$$

where L is the total number of locations that an ink droplet can land, $N(n)$ is the number of droplets in a given layer, and $C(n)$ is the number of possible arrangements for the particles in that layer. In section 2, it was determined that a resolution of 300^2 dots per square inch yields 5,025,780 total pixels per page (black or white), thus $L = 5{,}025{,}780$ locations. This function can be used to generate a set of values that denote the number of arrangements for each layer of the sector, $\{C(0), C(1), C(2), ...C(m)\}$. Also by assumption 11, although no two ink particles in a single layer can land in the same location, ink particles in subsequent layers *can*. Thus, the product of all items within the set of C values is the total number of possible arrangements for all ink particles in an entire sector. This quantity (hereafter referred to as Γ) can be represented in product notation as

$$\Gamma = \prod_{n=0}^{m} \binom{L}{N(n)} \tag{5}$$

where m continues to represent the number of layers in the sector.

6 Accounting for Multiple Correct Arrangements

Let us consider a page to be "successfully printed" when a particular subset of ΣN cells are filled with ink. Although there is exactly one subset of cells that results in a correctly printed page, those ΣN droplets of ink may be arranged in any manner within those cells. This repetition is illustrated in Figure 3 with a simple example of five ink droplets filling a single configuration of five target cells. These droplets are arranged into two layers, with three droplets in A and two in B. Two possible arrangements of these five droplets are shown in Figure 3. Both of the arrangements shown generate a "correctly printed" page despite being arranged differently. The total number of *correct* arrangements of these five particles (accounting for the fact that two particles within the same layer are identical) is described mathematically as a permutation with repetition, $\frac{5!}{3! \times 2!}$. This principle is applied to the sector as a whole by the following permutation with repetition:

$$\frac{\Sigma N!}{\prod_{n=0}^{m} N(n)!} \tag{6}$$

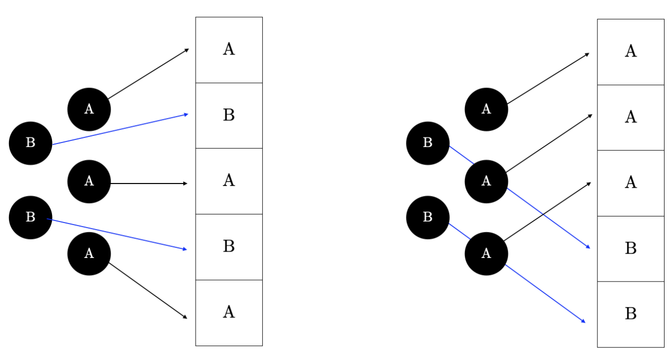

Figure 3: A simplified example of printing the letter "l" where five ink droplets are arranged into two layers, A and B. To successfully "print" this letter, all five target cells must be filled. In both the left and right scenarios all five cells are successfully filled, but both solutions are counted even though they both produce the same (single) outcome. To determine the number of correct arrangements, a permutation with repetition must be employed.

The actual probability of a page being printed is the ratio between the number of correct arrangements and the total number of arrangements. Γ provides the total number of arrangements; the number of correct arrangements is given by expression 6. The probability ratio, Λ is thereby expressed as

$$\Lambda = \frac{\Sigma N!}{\Gamma \prod_{n=0}^{m} N(n)!} \tag{7}$$

7 Computing Enormous Factorials

In order to evaluate Λ, factorials of L, ΣN, and various large $N(n)$ values must be computed. The largest factorial required is $L!$, which requires the computation of 5,025,780!. Exact computation of large numbers can take enormous amounts of computing time; a study of the number of legal moves in games of Go required 8000 central processing unit (CPU) hours to generate an exact number with an order of magnitude of 136 (for reference, 5,025,780! has an order of magnitude of 10^7) [17]. One commonly used function for calculating enormous factorials is Sterling's Approximation [18][19],

$$L! \sim \sqrt{2\pi L}(\frac{L}{e})^L \tag{8}$$

Although less computationally intensive than repeated multiplication, Sterling's Approximation for $L!$ requires the evaluation of $(\frac{L}{e})^L$, an operation that still requires a large amount of computational time. Furthermore, the resulting number (which also has an order of magnitude of 10^7) is too large to be stored by a conventional data type. Python 3.8.5 floating-point numbers are almost always mapped to IEEE-754 "double precision" [20]. The maximum value that Python can represent has an order of magnitude of 308, much too small to handle $(\frac{L}{e})^L$, and so Sterling's Approximation will not suffice for this computation. We can, however, effectively utilize Python's 17-digit floating point precision by exploiting the fact that multiplication in decimal space is equivalent to addition in logarithmic space. Thus,

$$L! = \prod_{n=0}^{L-1}(L-n) \tag{9}$$

$$\log_{10} L! = \sum_{n=0}^{L-1}\log_{10}(L-n) \tag{10}$$

$L = 5,025,780$, so its logarithmic representation is $\log_{10}(5,025,780) \approx 6.70$, which is a number easily handled by a floating point number. When evaluated in Python 3.8.5, the sum in equation 10 evaluates to the following:

$$\log_{10} L! \approx 31,496,109.622662853 \tag{11}$$

$$\therefore L! \approx 10^{31,496,109.622662853} \tag{12}$$

The power of ten in equation 12 can be factored into two terms: 10 raised to the integer portion of the power and 10 raised to the decimal portion of the power. 10 raised to the decimal portion can be easily evaluated by a computer, while 10 raised to the integer portion provides an order of magnitude for the computed decimal. This algorithm is illustrated by factoring the output of equation 12:

$$L! \approx 10^{0.622662853} \times 10^{31,496,109} \tag{13}$$

$$\approx 4.19 \times 10^{31,496,109} \tag{14}$$

Being that combinations are a function of factorials, this technique for computing large factorials can be used to create approximations for every factorial necessary to evaluate Λ.

8 The Probability of an Explosion Printing a Dictionary

With this technique for computing large factorials, the effective probability ratio, Λ_{ef}, that a single page of the dictionary is printed can be directly evaluated as expressed in equation 7. Thus,

$$\Lambda_{ef} \approx 1.78 \times 10^{-840,386}$$

However, being that a page can be printed either right side up or upside down and still be a recognizable page (and arbitrary rotations of the printed sphere are ignored par assumption 3), the actual probability ratio is twice Λ_{ef}, hence,

$$\Lambda \approx 3.56 \times 10^{-840,386}$$

This ratio can be translated into a more familiar form by taking its reciprocal; thus, the chance of a single page of the dictionary being successfully printed is 1 in $2.81 \times 10^{840,385}$. Raising Λ to the power of 973 (the total number of pages in the dictionary) will provide the probability that all pages are successfully printed. However, this power includes all possible arrangements of those pages. By assumption 12, the order of the printed pages does not matter, so the final probability ratio is given by Λ^{973} multiplied by the total number of arrangements of all printed pages. The total number of arrangements of all pages is given by the permutation ${}_{973}P_{973}$, as 973 positions can hold 973 unique pages. This permutation simplifies to 973!, and thus the final probability ratio, Ω, that a dictionary is printed is

$$\Omega \approx 973! \times \Lambda^{973} \quad (15)$$

$$\approx 1.30 \times 10^{-817,692,555} \quad (16)$$

Taking the reciprocal of Ω yields a more familiar form: 1 in $7.68 \times 10^{817,692,554}$.

9 Conclusions

The printing factory explosion thought experiment, up until this point, had not been formally evaluated. Although the original scenario—that of an explosion in a printing factory—is extremely complex to model in full, a highly simplified model consisting of an explosion of a single sphere of ink within a static paper shell can be mathematically modeled and estimated. Within our model, the probability of a dictionary being printed is found to be 1 in $7.68 \times 10^{817,692,554}$.

This probability can be compared to the probability of certain events described in popular media and other thought experiments. Students at the University of Leicester performed an investigation into the likelihood that a person moving at $0.99c$ could quantum tunnel through a solid wall, finding the likelihood of such an event to be $10^{-10^{28}}$[4]. On the other hand, the probability of a monkey randomly typing the complete works of Shakespeare has been estimated at about $10^{-10^{7.15}}$[5]. We calculated the probability of an explosion printing a dictionary under our model to be approximately $10^{-10^{9}}$, which is much higher than the probability of a person quantum tunneling through a solid wall and much lower than the probability of a monkey randomly typing the entire works of Shakespeare.

We present this as our estimate for the probability of a dictionary being printed in the circumstances of this model. It is important to note, however, that this does not establish an estimate for the probability of the entire print factory explosion thought experiment, as the sheer scale of an explosion in a printing factory introduces factors that could drive the likelihood of a dictionary being printed either up or down from that of our model. For instance, a print factory could be modeled as a series of many explosions identical to this one, effectively driving the likelihood up. Furthermore, this model assumes that the ink droplets must fill a specific arrangement perfectly to produce a dictionary, whereas it is much more likely to produce a dictionary that has almost all droplets in the correct arrangement, with some out of place; setting thresholds that accept these imperfect dictionaries would also drive the likelihood up. Conversely, adding complexity to the atomization of ink droplets or assuming non-uniform ink distribution throughout the spray would introduce variables that could supply excessive or insufficient ink to particular pages, effectively driving the likelihood of success down. This model thus serves as an entry point for more complex analyses of the thought experiment, opening the door for the evaluation of higher-fidelity models that can establish a more accurate estimate of the likelihood that a dictionary is printed in the explosion of a printing factory.

References

[1] A. Loeb, R. A. Batista, and D. Sloan. Relative likelihood for life as a function of cosmic time. *Journal of Cosmology and Astroparticle Physics*, 2016(08):040–040, Aug 2016.

[2] D. Kipping. An objective bayesian analysis of life's early start and our late arrival. *Proceedings of the National Academy of Sciences*, 117(22):11995–12003, 2020.

[3] C. Scharf and L. Cronin. Quantifying the origins of life on a planetary scale. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 2016.

[4] A. Alam, J. Campbell, J. Phelan, and B. Warne. The flash and quantum tunnelling. *Journal of Physics Special Topics*, 15(1), 2016.

[5] E. R. Weaver. An exercise in probability. *Journal of the Washington Academy of Sciences*, 1965.

[6] C. Kittel and H. Kroemer. *Thermal Physics*. W. H. Freeman and Company, 1980.

[7] D. Page. Return of the boltzmann brains. *Physical Review D*, 78, 12 2006.

[8] B. Beckers and P. Beckers. A general rule for disk and hemisphere partition into equal-area cells. *Computational Geometry*, 45(7):275 – 283, 2012.

[9] E. Sher, T. Bar-Kohany, and A. Rashkovan. Flash-boiling atomization. *Progress in Energy and Combustion Science*, 34(4):417–439, 2008.

[10] M. Xu, Y. Zhang, W. Zeng, G. Zhang, and M. Zhang. Flash boiling: Easy and better way to generate ideal sprays than the high injection pressure. *SAE International Journal of Fuels and Lubricants*, 6:137–148, 04 2013.

[11] S. J. Noronha, S. Z. Basheer, M. N. Vijay, A. Alnajjar, B.K. Sharma, and N. Singh. A comparative study of different printed documents to estimate the type of printer used. *Journal of Forensic Research*, pages 1–7, 2017.

[12] T. N. Pappas, C.-K. Dong, and D. L. Neuhoff. Measurement of printer parameters for model-based halftoning. *Journal of Electronic Imaging*, 2(3):193 – 204, 1993.

[13] M. Myllys, H. Häkkänen, J. Korppi-Tommola, K. Backfolk, P. Sirviö, and J. Timonen. X-ray microtomography and laser ablation in the analysis of ink distribution in coated paper. *Journal of Applied Physics*, 117(14):144902, 2015.

[14] G. R. North, J. B. Valdes, E. Ha, and S. S. P. Shen. The ground-truth problem for satellite estimates of rain rate. *Journal of Atmospheric and Oceanic Technology*, 1994.

[15] K. Eriksson, C. Johnson, and D. Estep. *Double Integrals. In: Applied Mathematics: Body and Soul*. Springer, Berlin, Heidelberg, 2004.

[16] D. Hughes-Hallett, A. M. Gleason, and W. G. McCallum. *Calculus: Single Variable*. John Wiley & Sons, Inc., 7 edition, 2017.

[17] J. Tromp and G. Farneback. Combinatorics of go. *Lecture Notes in Computer Science*, 4630, 2007.

[18] P. Diaconis and D. Freedman. An elementary proof of stirling's formula. *The American Mathematical Monthly*, 93(2):123–125, 1986.

[19] S. Eger. Stirling's approximation for central extended binomial coefficients. *The American Mathematical Monthly*, 121(4):344–349, 2014.

[20] Floating point arithmetic: Issues and limitations. *Python 3.8.11 Documentation*, 2021.