Cut a cake randomly n times. What is the expected number of pieces?

Dimitar Petrov - dimitar.sp@outlook.com, October 30, 2024

Abstract

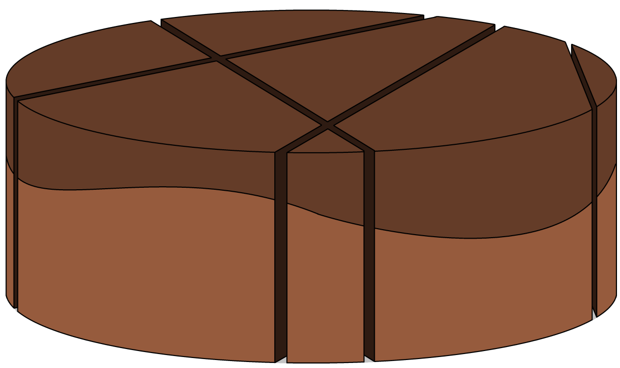
This study presents a geometric approach to a cake-cutting problem, focused on determining the expected number of pieces generated by n cuts. The main result, along with both upper and lower bounds for the total number of pieces, was derived analytically through induction on n. Additionally, a novel function was introduced to parameterize the cuts, allowing us to locate any intersections between cuts precisely. To validate the analytical models, a Python program was developed to generate meaningful statistics from large-scale simulations of cakes with varying numbers of cuts. The program was also used to propose a general probability distribution for the number of pieces formed by n cuts. This study seeks to consolidate and formalize existing solutions by approaching them within a consistent mathematical framework, leading to several open questions, such as the expected number of pieces after n cuts on a general convex shape, the effects of variable cutting inclinations, and the general probability distribution for the number of pieces generated by n cuts.

Keywords: Cake-cutting problem; Discrete geometry; Expectation; Minimum and maximum; Distribution; Intersections; Numerical Solution

1 Introduction

I first encountered this problem four months ago when Jane Street provided all the students, including myself, attending an international physics seminar with a shirt featuring the title of this study printed on the back − a challenge that, at first, no one seemed to take seriously. However, a month ago, during another physics seminar, a discussion around this problem arose and quickly took off. A problem that initially seemed impossible, had a solution within a week.

This study documents my independent work in a structured manner, seeking to formalize many of the existing solutions which were discovered during the write-up. While it does not introduce novel analytical solutions to the main results, it compiles scattered literature into a more cohesive, self-contained format. Furthermore, I provide evidence for the general probability distribution being a Gaussian, and introduce a parametrization which maps the cuts from "Cake space" into Cartesian space.



2 The Problem at Hand

Consider a perfectly cylindrical cake of arbitrary radius and height, which is to be cut n times along the vertical plane, not necessarily through the center. Initially, the cake starts as a single piece, and each successive cut will create at least one additional piece. Naturally, three interesting questions arise: (i) What is the minimum possible number of pieces generated after n cuts? (ii) What is the maximum number? and (iii) What is the expected number of pieces?

Before addressing these questions, it is helpful to introduce some fundamental intuition.

- We can think of random cuts as a geometric choice of two points on the cake's circumference, which are then connected by a straight line. Therefore, the circumference of the cake will be particularly relevant to the problem. This circumference is defined by the set $C = \{(x, y) \in \mathbb{R}^2 : x^2+y^2 = r^2, r \in \mathbb{R}^+\}$.
- The problem will be analyzed within the ZFC (Zermelo-Fraenkel with Choice) axiomatic framework, given the nature of point selections on the circumference and subsequent proofs for the existence of certain piece configurations.
- The height of the cake is irrelevant to the problem, as the cuts are fixed to the vertical plane, and their inclination is not a degree of freedom.
- The radius of the cake, r, is also not relevant! Indeed, for any $r \in \mathbb{R}$, the cardinality of C remains the same − an uncountable infinity, $\mathfrak{c} = |\mathrm{R}|$. This means that the number of choices available for cuts is unaffected by the radius. When expressing C in polar coordinates as $\{(r, \theta) : \theta \in [0, 2\pi]\}$, we note that θ is the only degree of freedom when choosing

the two points on the circumference. To this extent, a cut will be defined by the pair of angles at which it touches the circumference.

- Note that the cyclic nature of the circumference allows us to choose angles with $\theta > 2\pi$. All angles obey the relationship: $\theta = \theta_{\text{mod } 2\pi}$.

- Finally, the question of determining how many pieces are generated by n cuts is related to counting the number of intersections formed by n lines. Such intersections form depending on how the $2n$ points on the circumference – uniquely identified by the angle they subtend from the $\theta = 0$ axis – are connected by straight lines.

3 Solution

In this section, we will address the three previously posed questions **for a cake of unit radius**.

We begin by restating the vacuous fact that each cut necessarily creates a new piece by splitting the cake (ignoring any intersections). For those who may have trouble accepting this, recall that a piece is defined by connecting any two points $\theta_1, \theta_2 \in [0, 2\pi]$ with a straight line. This operation will partition the circumference – and thus the cake – into two disjoint segments, $[\theta_1, \theta_2] \cup (\theta_2, \theta_1 + 2\pi)$, thus increasing the piece count by one.

Moreover, each time an intersection between two cuts occurs, the number of pieces increases by one. This becomes evident when two cuts are examined in isolation: if they intersect, four pieces are created; if they do not, only three pieces result (refer to figure 3). It is important to keep in mind, and to avoid double-counting these intersections in the following sections!

This analysis allows us to assert that the total number of pieces generated by $n \geq 1$ cuts is given by

$$P(n) := 1 + \sum_{k=1}^{n} \left(\frac{1}{2} i_k + 1, \right),$$

and $P(0) := 1$. Where in the above expression, i_k the number of intersections formed along the k^{th} cut. Much like the handshaking lemma, the factor of $\frac{1}{2}$ accounts for the double-counting of intersections. As a last comment, we shall reduce the above expression to

$$P(n) = 1 + I(n) + n, \tag{1}$$

where $I(n) = \sum_{k=1}^{n} \frac{1}{2} i_k$ is the total number of intersections. In the following sections, i_k will be the subject of interest.

3.1 Minimum number of pieces, $P_{\text{min}}(n)$.

From expression (1), it is evident that $P(n)$ is minimized whenever $I(n)$ is also minimized. In the following argument, we will demonstrate the existence of a cut configuration which yields $I_{\text{min}}(n) = 0$, so that $P_{\text{min}}(n) = n+1$.

Proposition 1: *For n cuts, there exists a configuration for which the minimum number of intersections on each cut is $i_k = 0, \forall k \leq n$.*

Proof: To show this result by induction on $n \in \mathbb{N}$, take the base case $n = 2$ and consider an initial cut $\alpha_1, \alpha_2 \in [0, 2\pi]$. Since $\alpha_1 \neq \alpha_2$, then the arc-length $|[\alpha_1, \alpha_2]| \neq 0$, allowing us to place a second cut $\beta_1, \beta_2 \in (\alpha_1, \alpha_2)$, which does not intersect the first.

Suppose that n cuts can be formed in such a way without intersecting, with the n^{th} cut placed at $\theta_1, \theta_2 \in [0, 2\pi]$. Then, the $(n+1)^{\text{th}}$ cut can be freely chosen as $\phi_1, \phi_2 \in (\theta_1, \theta_2)$, thereby avoiding intersections entirely. Figure 1 shows how this cut configuration does not result in intersections. This shows that for all $k \leq n$, $i_k = 0 =: i_{\text{min}}$, as required.

■

Using equation (1), with $I_{\text{min}}(n) = 0$ from its definition, the minimum number of pieces generated by n cuts is given by:

$$P_{\text{min}}(n) = n + 1 \tag{2}$$

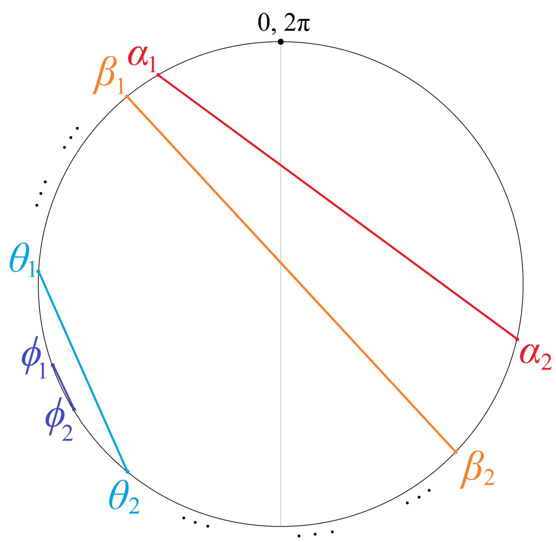


Figure 1: Inductive construction of the minimal cut configuration in Proposition 1.

3.2 Maximum number of pieces, $P_{\text{max}}(n)$.

The approach for determining the upper bound is not unique, but a similar argument to Proposition 1 will be made.

Proposition 2: *For n cuts, there exists a configuration for which the maximum number of intersections on each cut is $i_k = n - 1, \forall k \leq n$.*

Proof: As a base case with $n = 3$, select an arbitrary first cut $\alpha_1, \alpha_2 \in [0, 2\pi]$, and a second cut that intersects the first by choosing $\beta_1, \beta_2 \in [0, 2\pi]$ such that $\alpha_1 < \beta_1 < \alpha_2 < \beta_2$. Next, we select a third cut that intersects the first two without passing through their intersection. This can be achieved by choosing $\gamma_1 \in (\alpha_1, \beta_1)$ and $\gamma_2 \in (\alpha_2, \beta_2)$, as illustrated in figure 2. Each cut is intersected twice, as required.

[1]The method of avoiding intersections is understood better when cuts are parameterized. This is discussed in the appendix.

By induction, if the n^{th} cut is chosen appropriately with $\theta_1, \theta_2 \in [0, 2\pi]$, the $(n+1)^{\text{th}}$ cut can always be selected with $\phi_1 \in (\alpha_1, \theta_1)$ and $\phi_2 \in (\alpha_2, \theta_2)$, and without passing through any previous intersections [1].

To count the number of intersections due to n cuts, notice that with this construction, between every two points which define a cut there will be another $(n-1)$ points belonging to other cuts that will necessarily be intersected. Thus, for all $k \leq n$, $i_k = n - 1 =: i_{\max}$. ■

This implies that the total number of intersections is:

$$I_{\max}(n) = \sum_{k=1}^{n} \frac{1}{2} i_{\max} = \frac{1}{2} \sum_{k=1}^{n} (n-1)$$
$$\implies I_{\max}(n) = \frac{1}{2} n(n-1).$$

Finally, we derive an expression for the maximum number of pieces $P_{\max}(n)$ as follows:

$$P_{\max}(n) = 1 + I_{\max}(n) + n = 1 + \frac{1}{2} n(n-1) + n$$

which simplifies to:

$$P_{\max}(n) = \frac{1}{2}(n^2 + n + 2). \tag{3}$$

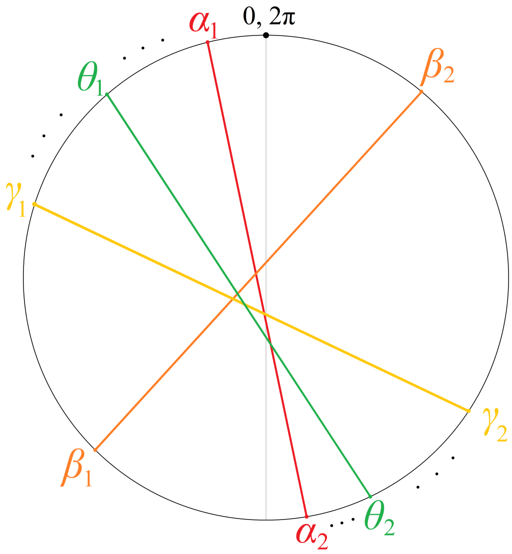


Figure 2: Inductive construction of the maximal cut configuration in Proposition 2.

3.3 Expected number of pieces, $\langle P(n) \rangle$.

We now come to the main result presented in this study.

Proposition 3: *For n cuts, the expected number of intersections on each cut is $\langle i_k \rangle = \frac{1}{3}(n-1), \forall k \leq n$.*
Proof: We will show the required result by induction on $n \in \mathbb{N}$. Take $n = 2$ as the base case, and focus on the number of ways two lines can be drawn connecting four points (refer to figure 3). It is clear that from the three possible cut configurations, only one will produce an intersection. Hence, $\langle i_k \rangle = \frac{1}{3} = \frac{1}{3}(2-1)$, for $k \in \{1, 2\}$.

As the inductive hypothesis, assume that for n cuts, $\langle i_k \rangle = \frac{1}{3}(n-1)$, $\forall k \leq n$. Consider now a cake with n cuts and an arbitrary $(n+1)^{\text{th}}$ cut. To count the new value of $\langle i_k \rangle$, notice that the new cut will once again have a $\frac{1}{3}$ chance of intersecting each previous cut. Alternatively, think about how the new cut will add another $\frac{1}{3}$ probability of an intersection occurring on each cut. Thus $\langle i_k \rangle = \frac{1}{3}(n-1) + \frac{1}{3} = \frac{1}{3}((n+1)-1)$ for all $k \leq (n+1)$, as required to complete the proof. ■

This shows that the average number of intersections in this case is:

$$\langle I(n) \rangle = \sum_{k=1}^{n} \frac{1}{2} \langle i_k \rangle = \frac{1}{2} \sum_{k=1}^{n} \frac{1}{3}(n-1)$$
$$\implies \langle I(n) \rangle = \frac{1}{6} n(n-1).$$

For the expected number of pieces, we obtain:

$$\langle P(n) \rangle = 1 + \langle I(n) \rangle + n = \frac{1}{6}(n^2 + 5n + 6). \tag{4}$$

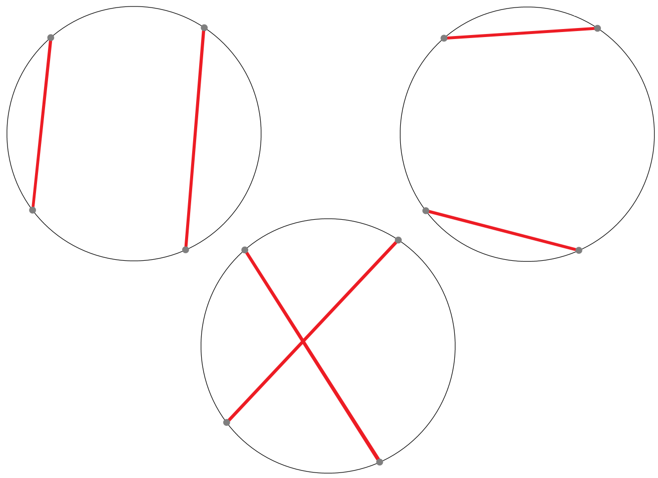


Figure 3: Base case $(n = 2)$ for Proposition 3 to obtain $\langle i_k \rangle$.

There is an opportunity to also derive this result through a recursive argument. For n cuts, there are $(2n-1)!!$ ways to join the $2n$ points on the circumference (Gould and Quaintance [2012]). Let $I_\Sigma(n)$ denote the sum of all intersections across all possible configurations. It can be shown that using:

$$\langle I(n) \rangle = \frac{I_\Sigma(n)}{(2n-1)!!},$$

one can derive the recurrence relation:

$$\langle I(n) \rangle = \langle I(n-1) \rangle + \frac{1}{6}(n-1), \langle I(0) \rangle = 0,$$

which when solved through conventional methods, yields an identical result.

4 Computer Simulation

A publicly accessible cake-cutting simulator is available on my GitHub profile [here]. This tool enables the random generation of any number of cuts on a cake and features an automatic piece-counting algorithm.

The documentation for this program will not be featured here as it is beyond the scope of this study, however it is a tool which helped me visualize the complexity of the problem more clearly. Actually, it was through this program that I first recognized the quadratic nature of $\langle P(n)\rangle$ which enabled the derivation of the analytic result. Furthermore, it gave insight into the Gaussian nature of the probability distribution of $P(n)$.

4.1 Validation of theoretical model.

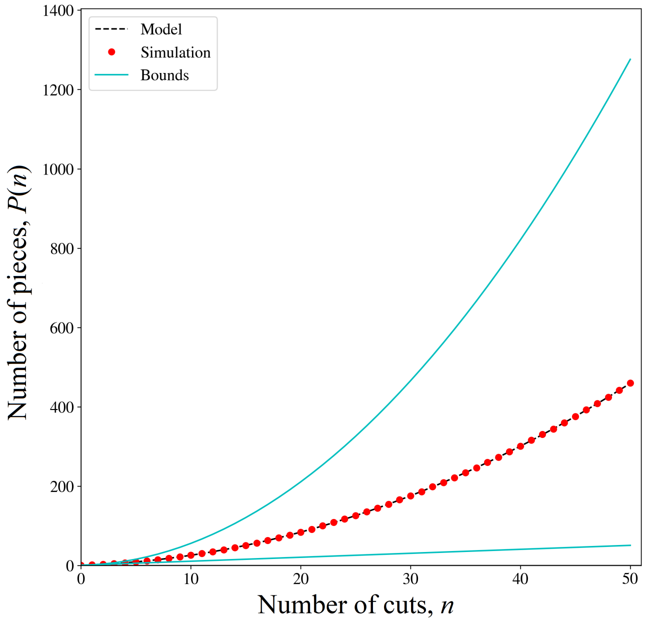


Figure 4: Typical results derived from the simulation. Each data point is obtained by taking the average number of pieces from 10,000 cakes simulated with a corresponding number of cuts.

To validate the theoretical expectation value in expression (4), a simulation of 1 million cakes in total was conducted. Figure 4 depicts the expected number of pieces for a cake with n cuts in red, alongside the theoretical upper and lower bounds in blue, and the theoretical expectation value from (4) in black. The polynomial of best fit for the simulated data in red was found to be $\langle P(n)\rangle_{\text{best fit}} = 0.1666n^2 + 0.8363n + 0.9413$, obtained using `np.polyfit()`, which closely aligns with the expanded form of equation (4), as was to be expected.

It was observed that for simulations involving cakes with more than 100 cuts, the polynomial estimate struggled to accurately determine the coefficient of n and the constant term, as they lose significance at larger scales. Conversely, for simulations with very few cuts, the polynomial fit encountered difficulties in accurately determining the coefficient of n^2. This issue could be resolved by simulating, for instance, ten times as many cakes for each n, which will take a significantly longer to compute.

However, this issue is largely irrelevant, as a simple visual comparison between the data points and the theoretical model consistently shows perfect alignment, making precise polynomial fitting unnecessary.

4.2 Numerical form of the distribution of $P(n)$.

For a fixed n, we have already determined the range of $P(n)$ in equations (2) and (3), as well as the expectation in (4). However, the general analytic form of the probability distribution of $P(n)$, denoted by $\mathbb{P}[P(n)]$, remains an open question.

As an initial approach to this problem, a simulation was conducted on 2 million cakes each with $n = 100$ randomly generated cuts. The probability distribution for the number of pieces observed is displayed in figure 5. The distribution exhibits a characteristic Gaussian shape, which motivated the fit of a Gaussian model of the form:

$$G(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{P(n)-\mu}{\sigma}\right)^2}$$

where μ is the mean, and σ is the standard deviation.

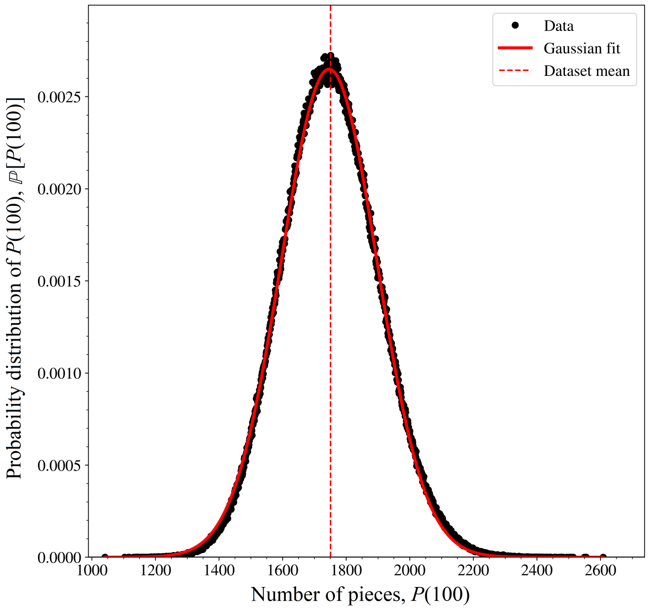


Figure 5: Probability for number of pieces formed by 100 cuts, $\mathbb{P}[P(100)]$, against number of pieces, $P(100)$. Data was obtained from a simulation of 2 million cakes with 100 cuts each.

The curve fitting program yielded the optimal parameters for fitting a Gaussian curve through the data points as: $\mu_{\text{best fit}} = 1,746.20 \pm 0.15$ and $\sigma_{\text{best fit}} = 150.33 \pm 0.12$. Comparing $\mu_{\text{best fit}}$ with the theoretic expectation derived in equation (4), $\mu_{\text{theory}} = 1,751$, shows 99.7% accuracy and 99.9% precision. Similarly, using the defining expression for the standard deviation:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n}(x_i - \mu)^2},$$

the standard deviation of the dataset was found to be 150.45, which when compared to that of the Gaussian fit, $\sigma_{\text{best fit}}$, yielded an accuracy of 99.9% and precision of 99.9%.

However, while theory predicts a minimum-to-maximum range of 101 to 5,051 pieces, the simulation produced a much narrower range of 1,113 to 2,541 pieces. This discrepancy is unsurprising, given the extremely strict conditions required to generate the theoretical boundaries, which are unlikely to occur even after 2 million attempts.

Lastly, even though the dataset appears to be a perfect Gaussian distribution, there seems to be a slight right-skew present in the regions away from the mean. This skew might be attributed to inherent inaccuracies either in the simulation or the curve-fitting algorithm. The peak is also not located at the midpoint between $P_{\min}(100)$ and $P_{\max}(100)$.

5 Conclusion and Future Work

In this study, we explored the nature of cuts made on ordinary circular cakes under highly idealized conditions. Future research could focus on several interesting extensions. First, the analytical form of the probability distribution of $P(n)$ remains an open question. Although the distribution appears to resemble a Gaussian due to the random nature of the cuts, a rigorous mathematical argument is needed to confirm this.

Several alternative models were also tested for their conformity to the data. Notably, given that the distribution is discrete, I also attempted to fit a Binomial distribution, $B(N, p)$, to the data with a probability parameter, $p = 1/3$, and number of trials, $N = P_{\max}(n)$. While the expectation of the distribution, Np, coincided with the analytical one, the standard deviation, $\sqrt{Np(1-p)}$, showed very poor conformity to the data. As a result, this model was ruled out, despite initially appearing suitable based on its assumptions.

Returning to possible extensions of this study, it would also be worth investigating the expected number of pieces after n cuts on more general convex bodies such as cuboids or even spheres. Additionally, studying the effects of a variable cut inclination will surely yield further productive discussions.

It should be noted that the idea of generating random cuts can be ambiguous, as is the case within Bertrand's paradox. A general discussion on the various methods used to produce random cuts can be found in the paper by Jevremovic and Obradovic [2011]. In this study, the problem of determining the expected number of pieces was approached by assuming that points on the circumference were selected based on a uniform distribution over the interval $[0, 2\pi]$. However, other piece-generating methods may yield different results. This approach was solely chosen based on my personal interpretation of the problem.

As a general comment on this study, I was unaware that this problem had already been addressed by Shaye [2024] until very recently. However, none of the available resources offered a formal and thorough analysis of the problem, often lacking the level of detail and rigor that this study aims to provide. I hope this work serves as a strong foundation for further discoveries in related problems.

For any additional comments or inquiries, feel free to contact me via email. Lastly, I would like to extend my thanks to Luca, Davide, and especially Giovanni, without whom this study would not have been possible.

6 Appendix

6.1 Parametrization of cuts.

In some cases, it is more convenient to represent the circumference of the cake as a straight-line interval $[0, 2\pi]$, with cuts depicted as sinusoidal-like curves originating from the angles that define them (see figure 6). Geometrically, this is done by disconnecting the circumference at $\theta = 0$, chosen arbitrarily, and flattening the circular boundary onto a linear axis.

The mapping from "Cake space" onto Cartesian space involves using the radial distance between each point on the cut and a corresponding point on the circumference. This is necessary to preserve the geometrical properties of the cuts, particularly for determining the location of intersections.

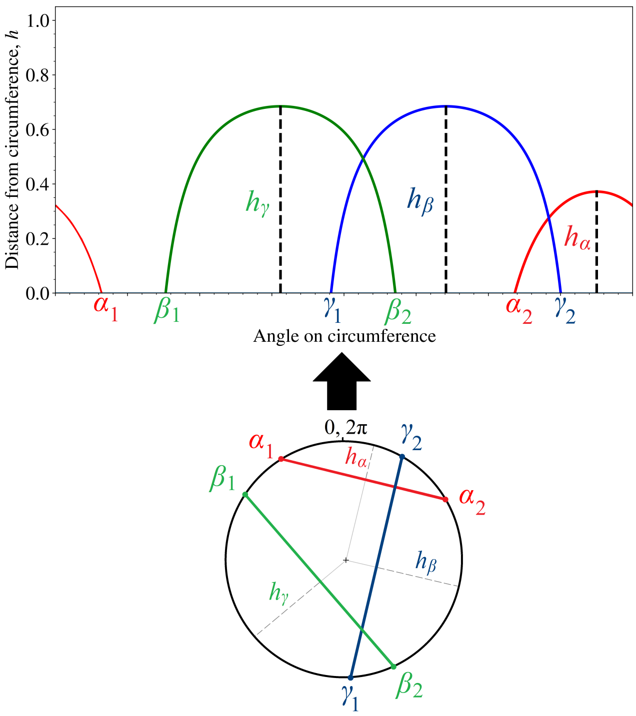


Figure 6: The one-to-one correspondence between cuts on a cake and their representation on a Cartesian plane.

The function that assigns a height, h_{α_1,α_2}, to a cut between the angles α_1 and α_2 is given by:

$$h_{\alpha_1,\alpha_2}(\theta) = 1 - \frac{\cos[\pi\chi + (-1)^\chi(\alpha_2 - \alpha_1)/2]}{\cos[\pi\chi + (-1)^\chi((\alpha_2 + \alpha_1)/2 - \theta)]} \quad (5)$$

where $\theta \in [\alpha_1, \alpha_2]$, and where $\chi := \left\lfloor \frac{\alpha_2 - \alpha_1}{\pi} \right\rfloor$, with $\lfloor \cdot \rfloor$ denoting the floor function. This ensures that the shortest path between the two angles is chosen automatically, and resolves the case when the cut passes over the periodic boundary $0 \equiv 2\pi$, which may otherwise cause a discontinuity. Note that whenever $2\pi > |\alpha_2 - \alpha_1| \geq \pi$, then $\chi = 1$, while $\chi = 0$ otherwise.

The procedure of mapping a cake onto Cartesian space begins with a radial line initially at $\theta = 0$, chosen arbitrarily. As the radial line rotates anticlockwise in the direction of increasing θ, the radial distance, $|P\theta|$, representing the distance between the point P, located at the intersection of a cut and the radial line, and corresponding point on the circumference, θ, is recorded (see figure 7). Completing a full 2π cycle allows the height of each cut to be measured over the interval $[0, 2\pi]$.

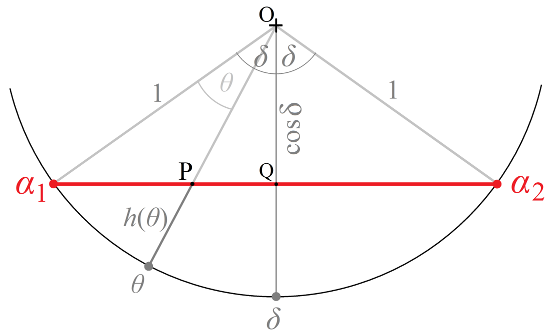


Figure 7: The construction required to derive the height of a cut using a radial line at $\theta \in [\alpha_1, \alpha_2]$. Point O denotes the center, point P the intersection between the radial line and the red cut, and the distance $|OQ|$ represents the vertically perpendicular height of the cut.

To derive the height function, let α_1, $\alpha_2 \in [0, 2\pi]$ be two angles which define a cut. The derivation can be broken down into the following arguments:

- The smallest angle separating the two points α_1 and α_2 is given by:

$$2\delta := 2\pi\chi + (-1)^\chi(\alpha_2 - \alpha_1).$$

  The χ term (defined above) is added to ensure that the separation is $2\delta \leq \pi$, while simultaneously describing the smaller of the two angles which can be used to label the separation between α_1 and α_2.

- In reference to figure 7, for a cake of unit radius, the vertically perpendicular distance, $|OQ|$, from the center of the cake to the cut is given by $\cos\delta$, which is derived through basic trigonometry.

- Next, the hypotenuse, $|OP|$, of triangle OPQ with angle $\angle POQ = \delta - (-1)^\chi(\theta - \alpha_1)$, can be found to be:

$$|OP| = \frac{\cos\delta}{\cos[\delta - (-1)^\chi(\theta - \alpha_1)]},$$

  where the angle $\angle POQ$ takes this interesting form due to the case when the shortest angle between the two endpoints traverses along the other side of the cake, as is the case with the red cut in figure 6. This necessitates the inclusion of the χ term once again. The parameter θ was translated by α_1 so that the domain of θ is $[\alpha_1, \alpha_2]$.

- Note that from our construction, $|OP| + |P\theta| = 1$. This allows for the height, $h_{\alpha_1,\alpha_2}(\theta) = |\theta P|$, to be expressed as:

$$h_{\alpha_1,\alpha_2}(\theta) = 1 - \frac{\cos\delta}{\cos[\delta - (-1)^\chi(\theta - \alpha_1)]}.$$

  The equivalence between this result and the simplified form in 5 can easily be verified.

Following this derivation, we can now analytically determine the location of an intersection between α_1, $\alpha_2 \in [0, 2\pi]$ and β_1, $\beta_2 \in [0, 2\pi]$, by solving $h_{\alpha_1,\alpha_2}(\theta) = h_{\beta_1,\beta_2}(\theta)$, for the unknown angle θ. This location will be given in terms of its angle, and its radial distance from the circumference.

With this in mind, placing a new cut that avoids passing through any intersections formed by previous cuts becomes a task of choosing two angles that generate a sinusoidal-like curve, which does not contain a set of discrete points, i.e. points arising from previous intersections. This framework aids to visualize the necessary conditions to prove proposition 2.

References

Henry Gould and Jocelyn Quaintance. Double fun with double factorials. *Mathematics Magazine*, 85(3):177–192, 2012. doi: 10.4169/math.mag.85.3.177. URL `https://doi.org/10.4169/math.mag.85.3.177`.

Vesna Jevremovic and Marko Obradovic. Bertrand's paradox: Is there anything else? *Quality & Quantity*, 46, 2011. doi: 10.1007/s11135-011-9553-7.

Jason Shaye. Piece(s) of cake?, 2024. URL `https://medium.com/technological-singularity/piece-s-of-cake-23c74ebc411f`.