Defining Randomness: Bertrand’s Paradox

Bilge Simav | 13 April 2026

1. The Random Chord Problem

In 1889, Joseph Bertrand posed a seemingly simple question about a random chord in a circle. Rather than showing off his creativity in math, the problem was intended to illustrate something deeper: that even in mathematics, the notion of randomness is not always as well-defined as it appears.

Consider an equilateral triangle inscribed in a circle, as shown in Figure 1. What Bertrand asked was: “Suppose a chord is chosen at random; what is the probability that the length of the chord is larger than a side of the triangle?”

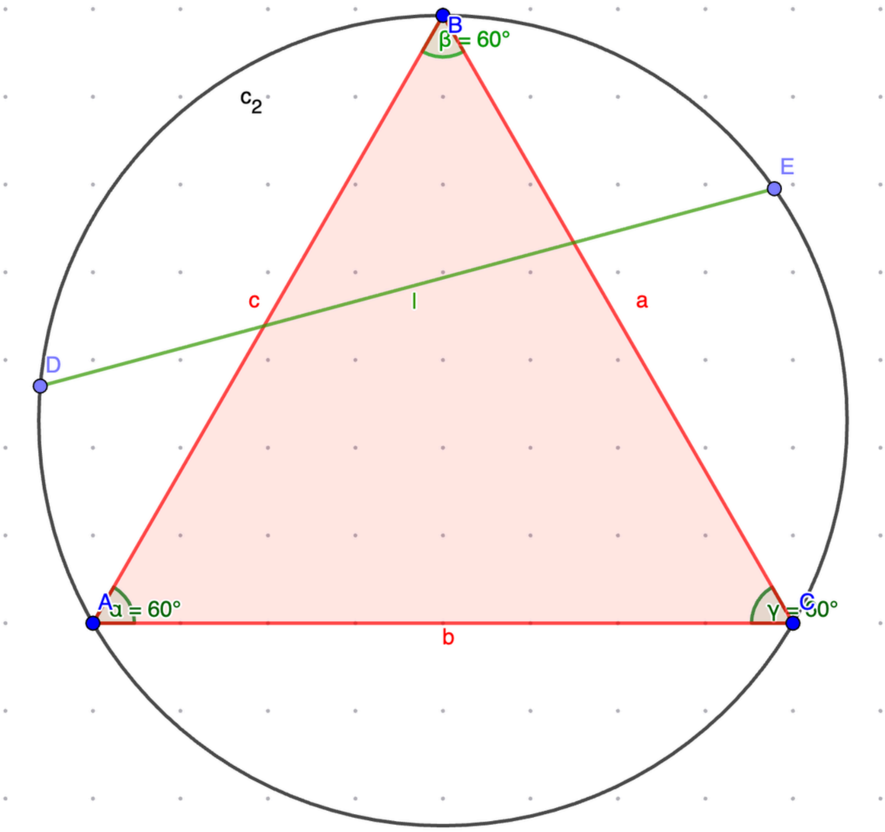

Fig. 1. A chord of length *l* with endpoints on the circle.

Bertrand suggested three ways to solve this problem. Each method defines randomness in its own way and yields a different probability.

METHOD 1: RANDOM ENDPOINTS

A chord is defined as the segment that connects 2 random points on the circle. However, when both endpoints are chosen freely, the problem becomes difficult to analyse geometrically, because we must consider all possible pairs of points on the circumference. Therefore, instead of working with both endpoints at once, we take advantage of the rotational symmetry of the circle to simplify the problem. Rotational symmetry means that the circle looks exactly the same after being rotated by any angle about its centre. In other words, there is no “special” point on the circumference: every point can be mapped to any other point by a rotation.

Now, suppose we fix a point A on the circle and consider all possible chords that start from A, formed by choosing a second endpoint uniformly at random. This gives us a distribution of chord lengths.

Next, take another point B on the circle. Because the circle is rotationally symmetric, we can rotate the entire circle so that point A moves exactly onto point B. Under this rotation, every chord starting from A is carried to a chord starting from B, and crucially, the lengths of the chords do not change under rotation. This means there is a one-to-one correspondence between the chords starting from A and the chords starting from B, such that corresponding chords have the same length. Therefore, the distribution of chord lengths from A must be identical to the distribution from B. Since this argument works for any pair of points on the circle, all points on the circumference are equivalent in terms of the chord length distribution.

Because of this, we are free to fix one endpoint (for convenience, say at the top of the circle) without changing the probability. With one endpoint fixed, the probability can now be determined by analysing the second endpoint's position on the circumference.

A chord will be longer than the side of the inscribed equilateral triangle if the second endpoint lies within a specific arc of the circle. In particular, this occurs when the endpoint falls on the arc opposite the fixed point that subtends the triangle’s side. Since this arc spans one-third of the circumference, the probability that the chord is longer than the triangle's side is **1/3.**

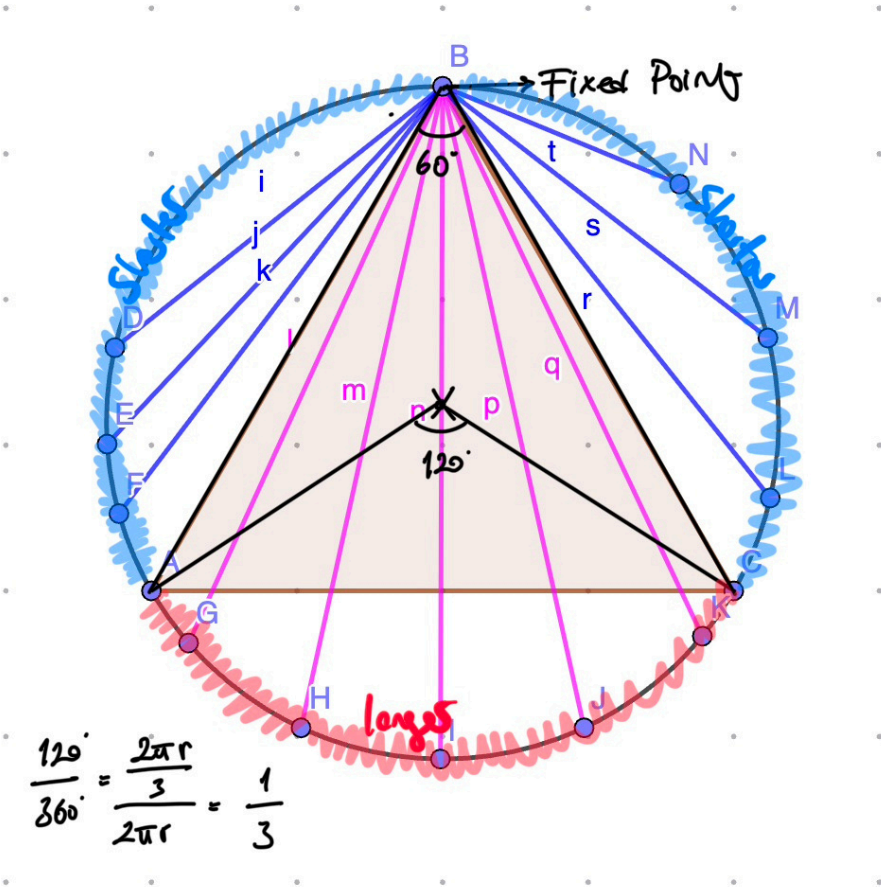

Fig. 2. End points of the chord uniformly distributed along the circumference of the circle.

In this method, we describe the position of the second endpoint using the angle θ, measured at the centre between the fixed point and the chosen point. Now consider two small angular intervals on the circle, say one of size dθ1 and another of size dθ2. Since the second endpoint is chosen uniformly along the circumference, there is no reason to prefer one interval over the other except for their size. Therefore, a larger interval must correspond to a higher probability of being selected. This means that the probability of selecting an interval depends only on its length, so the probability (dP) of selecting a small interval dθ is proportional to dθ.

$$dP = cd\theta \text{ for some constant } c \text{ since } dP \text{ is proportional to } d\theta$$

$$\int_0^{2\pi} dP = 1$$

$$\int_0^{2\pi} cd\theta = 1$$

$$c(2\pi) - c(0) = 1 \Longrightarrow c2\pi = 1 \Longrightarrow c = \frac{1}{2\pi}$$

$$dP = \frac{1}{2\pi} d\theta$$

METHOD 2: RANDOM RADIUS

The second method constructs a chord in a different way. First, we choose a direction from the centre, which determines a radius of the circle. Then, we select a point along this radius and draw a chord perpendicular to the radius at that point.

This point can lie anywhere between the centre (distance 0) and the edge of the circle (distance R). The key assumption is that this distance is chosen uniformly, meaning that every small segment of the radius has the same probability of being selected.

By rotational symmetry, points at the same distance r from the centre are equivalent.

To determine when a chord is longer than the side of the inscribed equilateral triangle, we compare this distance r to a critical value. From geometry, such a chord is longer than the triangle’s side if and only if its midpoint lies closer than R/2 to the centre.

Since r is chosen uniformly from the interval [0, R], the favourable outcomes correspond to $r \in [0, R/2]$, which is half of the total range. Therefore, the probability that a randomly chosen chord is longer than the triangle’s side is:

$$P = \frac{1}{2}$$

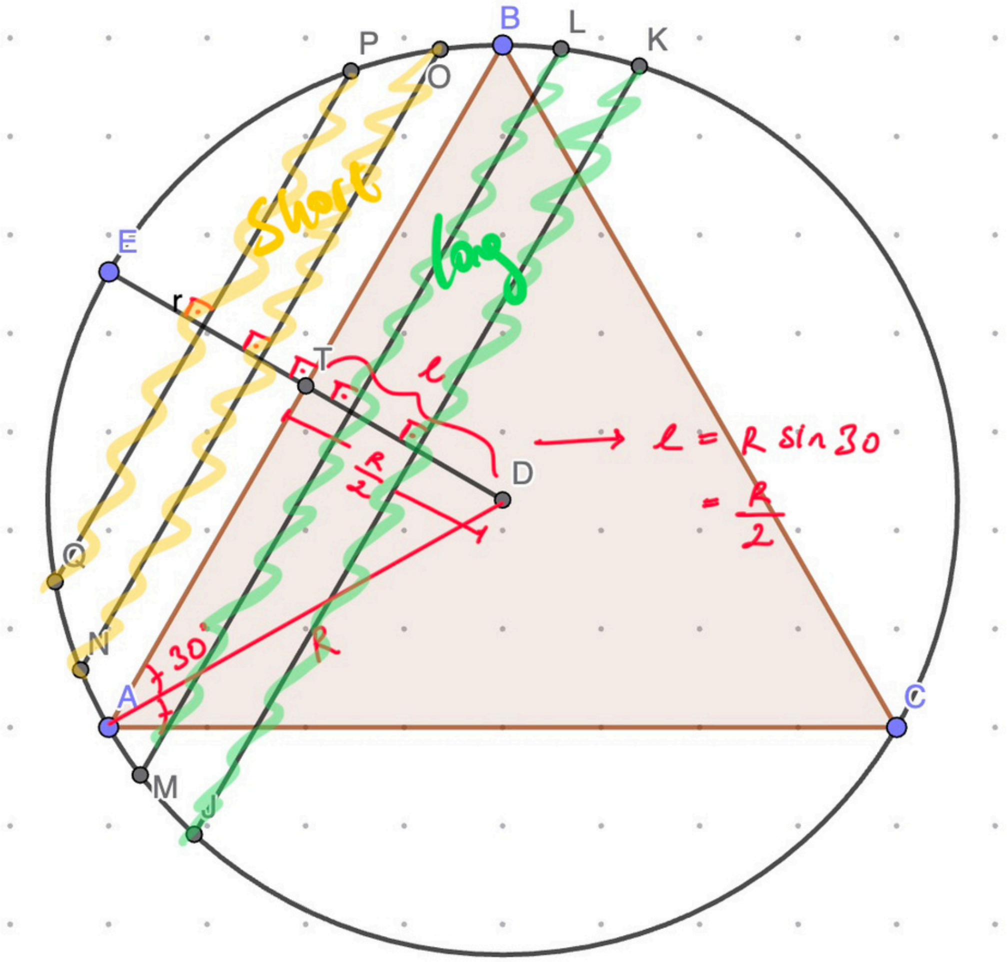

Fig. 3. Chords are constructed perpendicular to a randomly chosen radius, with their positions uniformly distributed along the radius.

In this method, the chord is determined by choosing a point along a radius, and its position is described by the distance r from the centre. By the same reasoning as method 1, the probability (dP) of selecting a small segment dr is proportional to dr:

$$dP = cdr \text{ for some constant } c \text{ since } dP \text{ is proportional to } dr$$

$$\int_0^R dP = 1 \text{ where R is the radius of the circle}$$

$$\int_0^R cdr = 1$$

$$c(R) - c(0) = 1 \implies cR = 1 \implies c = \frac{1}{R}$$

$$dP = \frac{1}{R}dr$$

METHOD 3: RANDOM MIDPOINT

The third method constructs a chord by first selecting a point uniformly at random from anywhere inside the circle and using it as the midpoint of the chord. A chord is then drawn such that this point lies at its centre.

The assumption of uniformity means that every region of equal area has the same probability of being selected.

By rotational symmetry, points at the same distance r from the centre are equivalent.

To determine whether a chord is longer than the side of the inscribed equilateral triangle, we again compare this distance r to a critical value. From geometry, such a chord is longer than the triangle's side if and only if its midpoint lies within a circle of radius R/2, centred at the origin.

Since points are chosen uniformly over the entire area of the circle, the probability is proportional to the area. The favourable outcomes correspond to the area of a smaller circle of radius R/2, while the total possible outcomes correspond to the area of the full circle of radius R. Therefore, the probability is given by the ratio of these areas:

$$P = \frac{\pi (R/2)^2}{\pi R^2} = \frac{1}{4}$$

For any point inside the circle (except the centre), there is a unique chord that has this point as its midpoint. When the midpoint lies at the centre, there are infinitely many possible chords (all diameters). However, this ambiguity is resolved by implicitly choosing a direction uniformly at random. Moreover, since the probability of selecting exactly the centre is zero in a continuous distribution, this degenerate case does not affect the overall probability.

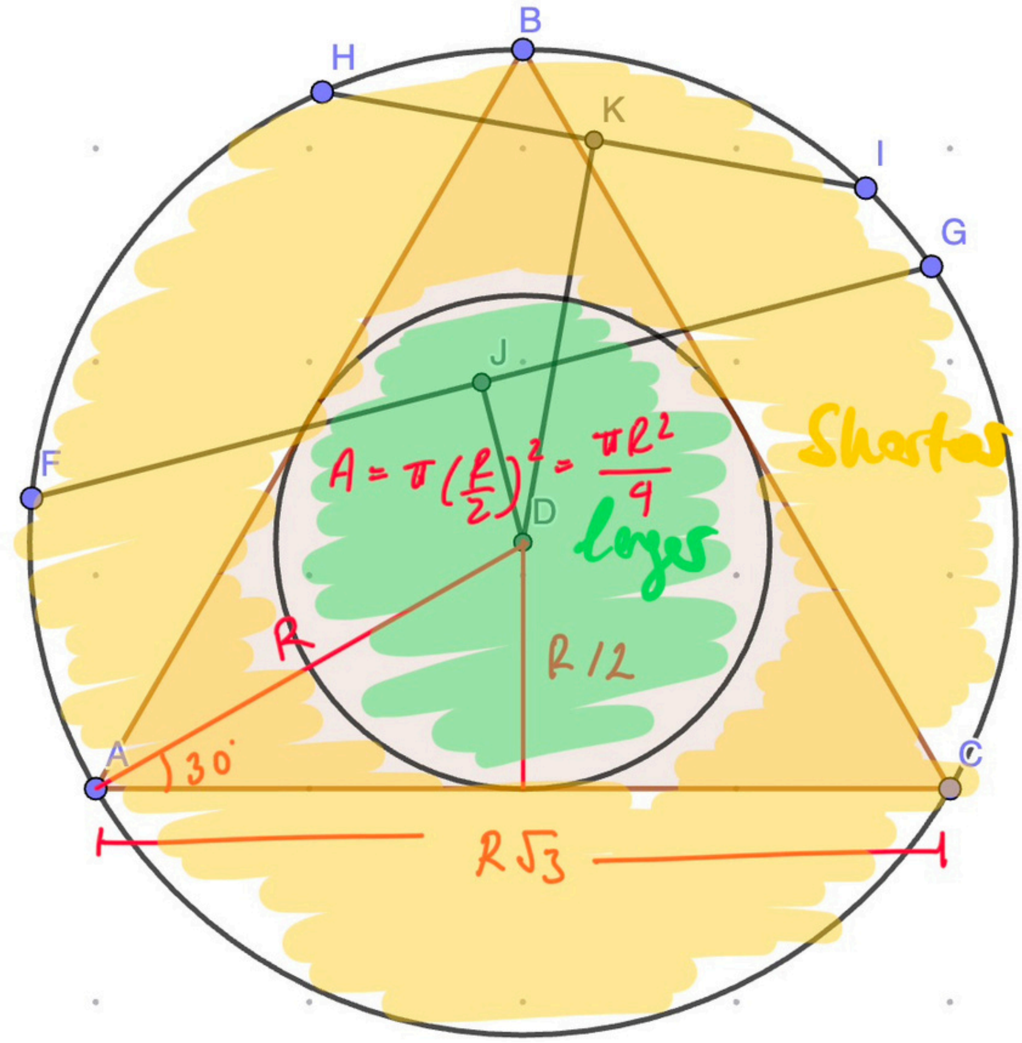

Fig. 4. Chords are constructed by uniformly selecting midpoints at random within the circle and drawing chords through them.

In this method, the midpoint of the chord is chosen from anywhere inside the circle, and its position is determined by its location in the plane. By the same reasoning explained in method 1, the probability (dP) of selecting a small region dA is proportional to dA:

$$\int_{circle} dP = 1$$

$$\int\int_{circle} c\,dA = c\int\int_{circle} dA = 1$$

$$c\pi R^2 = 1 \Rightarrow c = \frac{1}{\pi R^2}$$

$$dP = \frac{1}{\pi R^2} dA$$

2. Historical Solutions

CLASSICAL PERSPECTIVE:

The classical interpretation of probability assumes that outcomes are equally likely, but in this problem, it is unclear what should be treated as the fundamental outcome: points on the circumference, distances from the centre, or points within the area. As a result, the problem is not well-defined and does not admit a unique solution. The three solutions proposed by Bertrand are all valid within their own assumptions, and from a purely classical standpoint, there is no inherent reason to prefer one over the others.

EDWIN JAYNES' SOLUTION:

> **If the random process is not clearly defined, no additional assumptions should be imposed beyond the information given.**

In 1973, Edwin Jaynes proposed a resolution to Bertrand's paradox based on the principle of "maximum ignorance," which aligns with the principle of maximum entropy. He argued that, in the absence of additional information, one should not introduce any assumptions that are not explicitly given in the problem, which, in this case, correspond to assuming uniformity along the circumference (Method 1), uniformity along the radius (Method 2), or uniformity over the area of the circle (Method 3).

Jaynes observed that the problem does not specify the position or size of the circle, and therefore any valid probability distribution should remain unchanged under transformations such as translation and scaling. In other words, the solution should be invariant under shifts in position and changes in size.

Jaynes suggested expressing dP in terms of dL. The calculations below are not taken directly from his work, but are derived here using his framework.

Applying Jaynes' Framework: Method 1

Method 1 assumes that the angle θ is uniformly distributed.

$$dP = \frac{1}{2\pi} d\theta$$

Expressing the chord length in terms of the angle θ (the proof is illustrated in Figure 5).

$$L = 2R\sin\left(\frac{\theta}{2}\right)$$

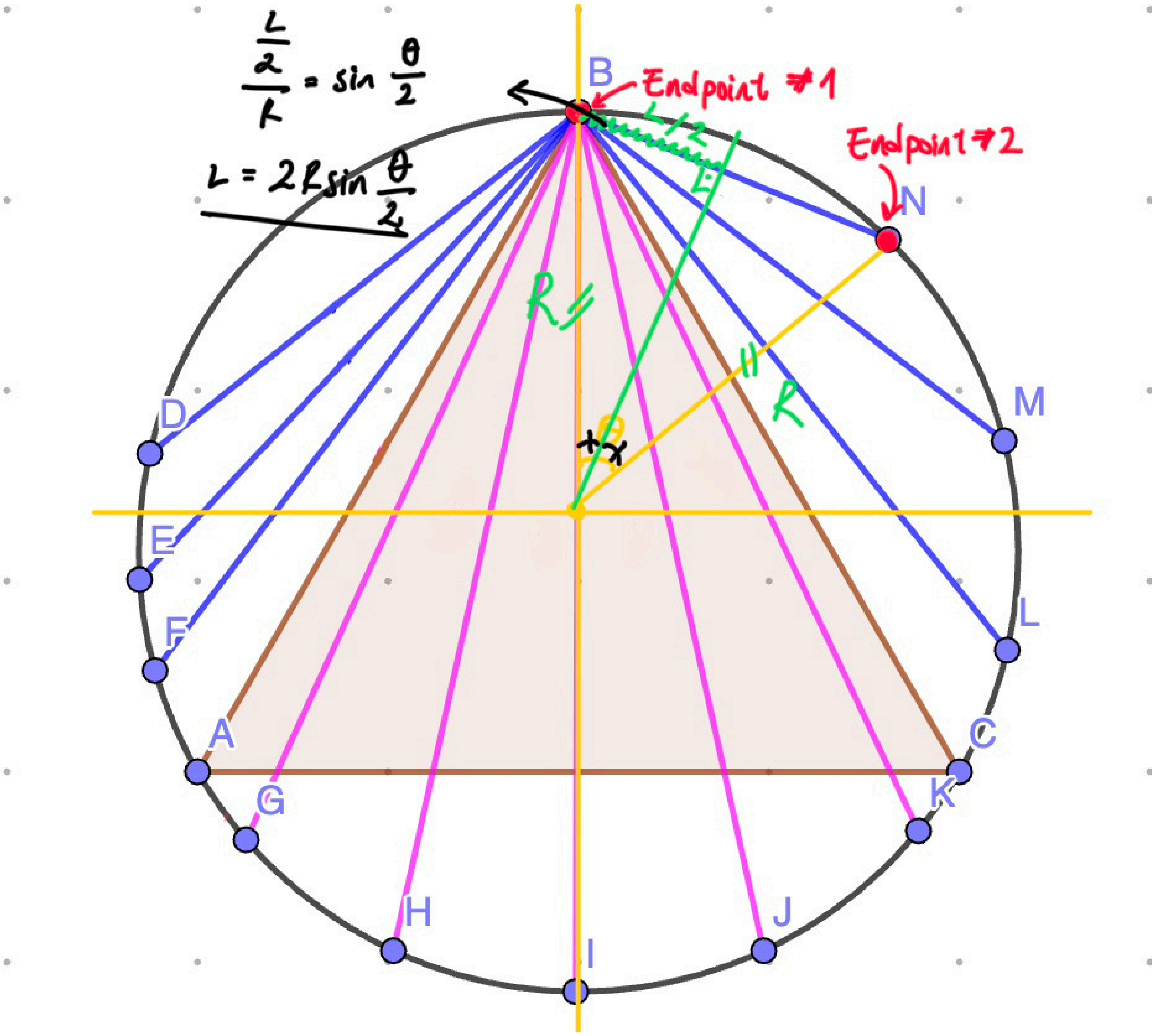

Fig. 5. The length of a chord.

Writing an expression for the probability using chord length for some function *f(L)*.

$$dP = f(L)\,dL$$

$$dP = \frac{1}{2\pi}d\theta = f(L)\,dL$$

Divide both sides by dL.

$$f(L) = \frac{1}{2\pi}\frac{d\theta}{dL}$$

Because dθ/dL can be negative, but probabilities can not, we'll take its magnitude.

$$f(L) = \frac{1}{2\pi}\left|\frac{d\theta}{dL}\right|$$

Having already derived an expression for L, we aim to find dL/dθ to later find dθ/dL, to substitute above.

$$L = 2R\sin\left(\frac{\theta}{2}\right)$$

Differentiate:

$$\frac{dL}{d\theta} = R\cos\left(\frac{\theta}{2}\right)$$

So:

$$\frac{d\theta}{dL} = \frac{1}{R\cos(\theta/2)}$$

Substituting for f(L)

$$f(L) = \frac{1}{2\pi}\left|\frac{1}{Rcos(\theta/2)}\right|$$

Rearranging the expression for the chord length to substitute for cos(θ/2):

$$L = 2R\sin\left(\frac{\theta}{2}\right)$$

$$\sin\left(\frac{\theta}{2}\right) = \frac{L}{2R}$$

$$\cos\left(\frac{\theta}{2}\right) = \sqrt{1 - \frac{L^2}{4R^2}}$$

Final Result:

$$f(L) = \frac{1}{2\pi R\sqrt{1-\frac{L^2}{4R^2}}}$$

$$dP = \frac{1}{2\pi R\sqrt{1-\frac{L^2}{4R^2}}}\,dL$$

Using the derived equation above to test the scaling invariance argument. λ is the scaling factor.

$$R \to \lambda R, \quad L \to \lambda L$$

Then:

$$f_{\text{new}}(L) = \frac{1}{2\pi(\lambda R)\sqrt{1-\frac{L^2}{4\lambda^2 R^2}}}$$

$$dP = f_{\text{new}}(L)\,dL$$

Conclusion:

$$f_{\text{new}}(L) \neq f(L)$$

The distribution changes with scaling. Method 1 is not scale invariant.

Applying Jaynes' Framework: Method 3

Method 3 assumes that the midpoint of each chord is chosen from a uniform distribution over the entire area of the circle.

$$dP = \frac{1}{\pi R^2}dA$$

Rewriting the area element dA using the polar coordinates

$$dA = r\,dr\,d\theta$$

Therefore:

$$dP = \frac{1}{\pi R^2}\,r\,dr\,d\theta$$

The angle θ is in the interval from 0 to 2π.

$$\int_0^{2\pi} d\theta = 2\pi$$

Using the identity above, integrating over θ:

$$dP = \frac{1}{\pi R^2} \cdot r \cdot (2\pi)\, dr$$

$$dP = \frac{2r}{R^2} dr$$

So the radial distribution is:

$$f(r) = \frac{2r}{R^2}$$

The probability in terms of chord length is:

$$dP = f(L)\, dL$$

And the length of the chord based on the given parameters is (proof is in Figure 6).

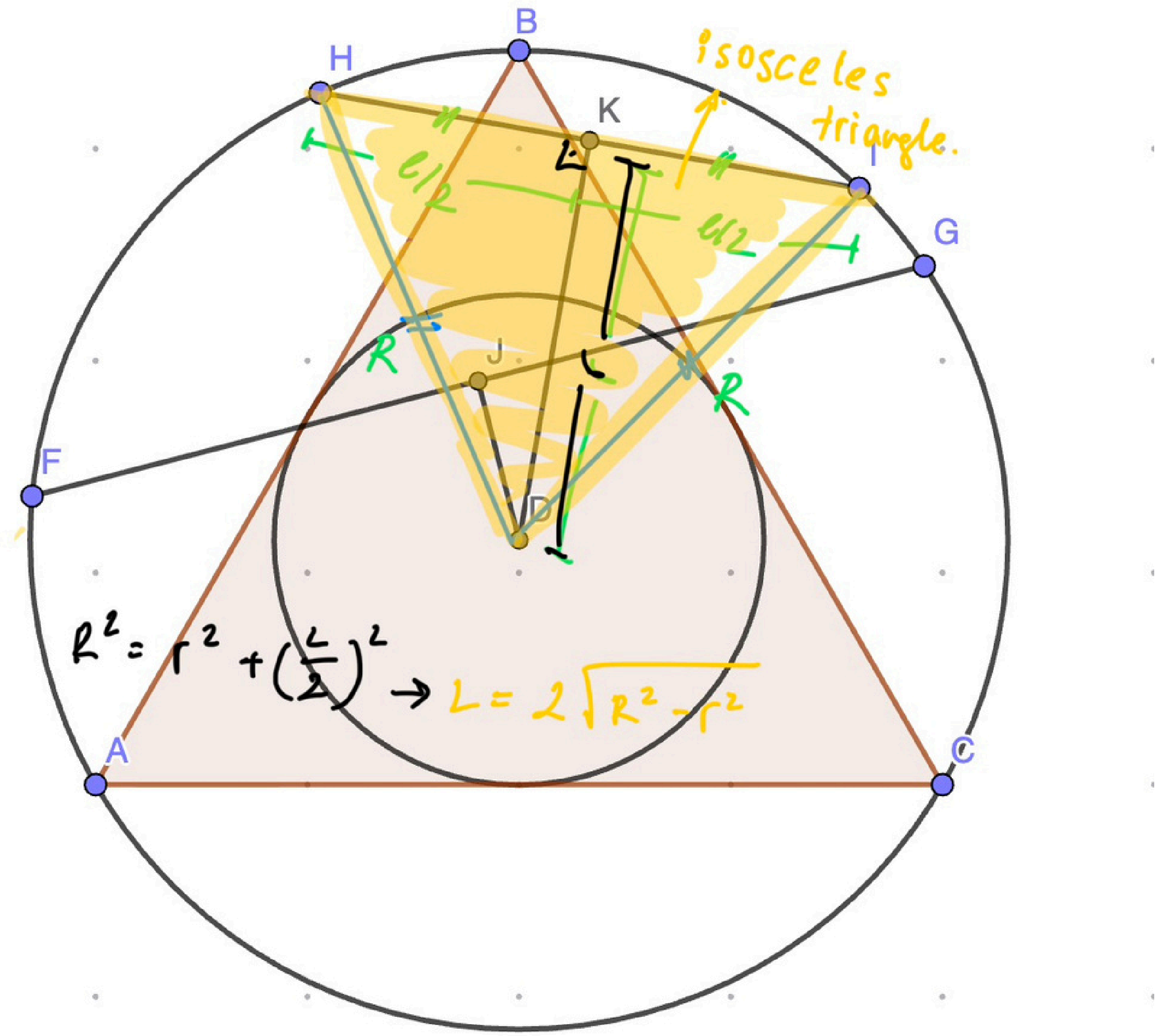

Fig. 6. The length of a chord.

$$L = 2\sqrt{R^2 - r^2}$$

Differentiate:

$$\frac{dL}{dr} = \frac{-2r}{\sqrt{R^2 - r^2}}$$

Because dL/dr can be negative, but probabilities can not, we’ll take its magnitude.

$$\left|\frac{dr}{dL}\right| = \frac{\sqrt{R^2 - r^2}}{2r}$$

The logic of the statement below is the same as the method for changing the variable for Method 1 (see JAYNES ON METHOD 1 section for detailed explanation)

$$f(L) = f(r)\left|\frac{dr}{dL}\right|$$

Substitute:

$$f(L) = \frac{2r}{R^2} \cdot \frac{\sqrt{R^2 - r^2}}{2r}$$

$$f(L) = \frac{\sqrt{R^2 - r^2}}{R^2}$$

Recall

$$L = 2\sqrt{R^2 - r^2}$$

$$\sqrt{R^2 - r^2} = \frac{L}{2}$$

Final Result:

$$f(L) = \frac{L}{2R^2}$$

Using the derived equation above to test the scaling invariance argument. λ is the scaling factor.

$$R \to \lambda R$$

$$f_{\text{new}}(L) = \frac{L}{2\lambda^2 R^2}$$

Conclusion:

$$f_{\text{new}}(L) \neq f(L)$$

The distribution changes with scaling. Method 3 is not scale invariant.

Applying Jaynes' Framework: Method 2

Method 2 assumes that the distance *r* of the chord from the centre is uniformly distributed along a randomly chosen radius.

$$dP = \frac{1}{R} dr$$

We want

$$dP = f(L)\, dL$$

The logic of the statement below is the same as the method for changing the variable for Method 1 (see JAYNES ON METHOD 1 section for detailed explanation)

$$f(L) = \frac{1}{R} \left| \frac{dr}{dL} \right|$$

Having proved the length of the chord expression for the analysis method 3, differentiating the expression.

$$L = 2\sqrt{R^2 - r^2}$$

$$\frac{dL}{dr} = \frac{-2r}{\sqrt{R^2 - r^2}}$$

Taking the absolute value

$$\left| \frac{dr}{dL} \right| = \frac{\sqrt{R^2 - r^2}}{2r}$$

Substituting:

$$f(L) = \frac{1}{R} \cdot \frac{\sqrt{R^2 - r^2}}{2r}$$

Rewriting the chord length expression in 2 different ways to substitute in the equation above

$$r = \sqrt{R^2 - \frac{L^2}{4}}$$

$$\sqrt{R^2 - r^2} = \frac{L}{2}$$

Final Result:

$$f(L) = \frac{L}{4R\sqrt{R^2 - \frac{L^2}{4}}}$$

Using the derived equation above to test the scaling invariance argument. λ is the scaling factor.

$$R \to \lambda R, \quad L \to \lambda L$$

$$f_{\text{new}}(L) = \frac{L}{4\lambda R\sqrt{\lambda^2 R^2 - \frac{L^2}{4}}}$$

Factor out λ inside the square root expression in the denominator:

$$\sqrt{\lambda^2 R^2 - \frac{L^2}{4}} = \lambda\sqrt{R^2 - \frac{L^2}{4\lambda^2}}$$

Recall f(L) and substitute:

$$f_{\text{new}}(L) = \frac{1}{\lambda} f\left(\frac{L}{\lambda}\right)$$

Using the dP=f(L)dL argument above:

$$P(a \le L \le b) = \int_a^b f(L)\, dL$$

Apply scaling and the interval [a, b] becomes [λa, λb]:

$$L' = \lambda L$$

Writing the new probability:

$$P_{\text{new}}(\lambda a \le L' \le \lambda b) = \int_{\lambda a}^{\lambda b} f_{\text{new}}(L')\, dL'$$

Substitute back:

$$L' = \lambda L \quad \Rightarrow \quad dL' = \lambda\, dL$$

$$\int_{\lambda a}^{\lambda b} f_{\text{new}}(L')\, dL' = \int_a^b f_{\text{new}}(\lambda L)\, \lambda\, dL$$

Using the scaled form derived before:

$$f_{\text{new}}(L) = \frac{1}{\lambda} f\left(\frac{L}{\lambda}\right)$$

So:

$$f_{\text{new}}(\lambda L) = \frac{1}{\lambda} f(L)$$

Substitute using L’=λL

$$\int_a^b f_{\text{new}}(\lambda L)\,\lambda\,dL = \int_a^b \frac{1}{\lambda} f(L)\cdot\lambda\,dL$$
$$= \int_a^b f(L)\,dL$$

Final Result:

$$P_{\text{new}}(\lambda a \le L' \le \lambda b) = P(a \le L \le b)$$

This means that probabilities are preserved under scaling. The scaling invariance condition is satisfied.

Therefore, by requiring invariance under translation and scaling and avoiding the introduction of additional assumptions, Jaynes’ framework uniquely identifies Method 2 as the only valid definition of randomness in this context.

4. Conclusion

I would love to finish my essay by writing about the maximum entropy principle, but due to word count constraints, I have to finish my essay here. I find this paradox interesting because it questions the philosophical interpretation of the idea of assigning probabilities to certain events. This paradox shows that probability depends not only on calculation, but on how randomness is defined.

Works Cited

Bertrand Paradox (Probability). Oslo University, 2021. *Oslo University*, www.uio.no/studier/emner/matnat/math/MAT4010/data/forelesningsnotater/ w-bertrand-paradox-(probability).pdf. Accessed 13 Apr. 2026.

T. M., and Joy A. Thomas. "Maximum Entropy." *Elements of Information Theory*, 2nd ed., Wiley-Interscience, 2006, pp. 409-25. Elements of Information Theory. *Wiley Online Library*, https://doi.org/10.1002/ 047174882x.ch12. Accessed 13 Apr. 2026.

Jaynes, Edwin T. *Probability Theory: The Logic of Science*. 2004 ed., Cambridge UP, 2022. Accessed 13 Apr. 2026.

Jaynes, Edwin T. "The Well-posed Problem." *Foundations of Physics*, vol. 3, no. 4, Dec. 1973, pp. 477-92. *Springer Nature*, https://doi.org/10.1007/bf00709116. Accessed 13 Apr. 2026.

"More on Bertrand's Paradox (with 3blue1brown) - Numberphile." *YouTube*, uploaded by Numberphile2, 21 Dec. 2021, www.youtube.com/watch?v=pJyKM-7IgAU. Accessed 13 Apr. 2026.

"The Principle of Maximum Entropy." *YouTube*, uploaded by Mutual Information, 6 July 2021, www.youtube.com/watch?v=2gTrsLVnp9c. Accessed 13 Apr. 2026.