Why Your Friends Are More Popular Than You (Statistically Speaking)

Joseph K and Jordan L

April 2025

Abstract

Probability theory began as a tool for analyzing games of chance in the 17th century, with foundational work by Blaise Pascal and Pierre de Fermat. Over time, it evolved into a powerful framework for understanding uncertainty across disciplines—from physics to social science. As the theory matured, it revealed that even simple averages can produce surprisingly counterintuitive results. One such example is the friendship paradox.

Here's a fun fact: your friends probably have more friends than you. This paper explores the mathematics behind that claim and shows how the paradox arises not from social failure, but from the logical structure of probability and sampling. Through this lens, we uncover how mathematical reasoning can explain everyday illusions.

1 History of probability

Probability started in the 1600s when people were studying gambling and insurance, and today it's an important tool used in both social sciences and natural sciences. The modern math of chance started in 1654 when French mathematicians Pierre de Fermat and Blaise Pascal wrote letters to each other, trying to solve a gambling problem from a gambler named the chevalier de Méré. Fermat used probabilities and Pascal used what we now call "expectation" to figure out how much money each player should get if the game was stopped early, and they both got the same answer. Even though an Italian named Girolamo Cardano worked on similar ideas over 100 years earlier, probability theory only became popular during the 17th-century scientific revolution when people started trusting math more.

2 Introduction

Imagine you're scrolling through social media. Everyone seems to be out with friends, while your weekend plans involve sitting in bed and maybe a cat video. Is everyone more connected than you — or is it just you?

Surprisingly, math has an answer: yes, your friends likely are more popular than you. But it's not your fault — it's statistically inevitable. This is the friendship paradox, and this paper explores why it happens and what it reveals about averages and probability.

3 Basic Probability Concepts

Before diving into the paradox itself, it is important to review some key ideas in probability theory that help explain why such counterintuitive results arise.

4 The basic principle of counting

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are mn possible outcomes of the two experiments.

4.1 Proof of the Basic Principle:

(1, 1), (1,2), ..., (1,n) (2, 1), (2,2), ..., (2,n)

(m,1), (m,2),..., (m,n).

We say the outcome is (i,j) if the first experiment give its ith result, and the second experiment then gives its jth result. So, the full set of possible outcomes forms a table with m rows(for the m outcomes of the first experiment), and each row has n elements(for the n outcomes of the second experiment). This shows the total number of outcomes $m \times n$.

4.2 Example:

A small community consists of 10 females, each of whom has 3 children. If one female and one of her children are to be chosen as parent and child of the year, how many different choices are possible?

Solution: By regarding the choice of the first female as the outcome of the first experiment and the following choice of one of her children as the second experiment, we can use the basic principle to understand that there are $10 \times 3 = 30$ possible outcomes.

5 Permutations

How many different ordered arrangements can be made from the letters **a**, **b**, and **c**? Each arrangement is a permutation.

If we list them out, we get: `abc`, `acb`, `bac`, `bca`, `cab`, and `cba` — a total of 6 arrangements.

We can also find this using the basic counting principle:

- There are 3 choices for the first letter,
- 2 choices for the second letter (since one is already used),
- and 1 choice left for the third letter.

So, the total number of permutations is:

$$3 \times 2 \times 1 = 6$$

General Case: Permutations of n Objects

Suppose now that we have n objects. Using the same reasoning as we did for the 3 letters, we can find the total number of different ordered arrangements (permutations) of these n objects.

For the first position, there are n choices. Once that is chosen, there are $n-1$ choices left for the second position, then $n-2$ for the third, and so on, until only 1 choice remains.

So, the total number of permutations is:

$$n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1 = n!$$

5.1 Example

$n!$ (read as "n factorial") is defined to equal $1 \cdot 2 \cdot \cdots \cdot n$ when n is a positive integer.

How many different batting orders are possible for a baseball team consisting of 9 players?

Solution: There are $9! = 362{,}880$ possible batting orders.

Combinations: Choosing r Items from n

We are often interested in how many different groups of r objects can be formed from a total of n objects. This is known as a **combination**.

Example

Suppose we want to choose a group of 3 items from the 5 items: A, B, C, D, and E.

First, let's count the number of ways to choose 3 items when the order *does* matter (this is a permutation):

$$5 \times 4 \times 3 = 60$$

However, since order doesn't matter in a group, each group of 3 is counted multiple times — exactly $3! = 6$ times (e.g., ABC, ACB, BAC, BCA, CAB, CBA).

So, to get the number of distinct groups (where order doesn't matter), we divide by 3!:

$$\frac{5 \times 4 \times 3}{3 \times 2 \times 1} = \frac{60}{6} = 10$$

So, there are 10 different groups of 3 that can be formed from 5 items.

General Rule

In general, the number of combinations of r items from a set of n items is given by:

$$\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!}$$

This counts all ways to choose r items from n, where order does not matter.

Notation

We write this number using the binomial coefficient notation:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

This is read as "n choose r." It represents the number of combinations (unordered groups) of r items chosen from n.

Special Cases

- $\binom{n}{n} = \binom{n}{0} = 1$: There is only 1 way to choose all n items, and only 1 way to choose none.
- If $r > n$ or $r < 0$, we define $\binom{n}{r} = 0$, since it's not possible to choose more items than we have or a negative number of items.

5.2 Sample Space and Events

A **sample space**, usually denoted by Ω, is the set of all possible outcomes of an experiment. An **event** is a subset of the sample space. For example, if we randomly pick a person from a group, the sample space might be all individuals, and an event could be "the person has more than 5 friends."

5.3 Axioms of Probability

Probability is based on three fundamental rules:

1. **Non-negativity**: For any event A, $P(A) \geq 0$.
2. **Normalization**: The probability of the entire sample space is 1, so $P(\Omega) = 1$.
3. **Additivity**: If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

5.4 Random Variables

A **random variable** is a function that assigns a number to each outcome in the sample space. For instance, we might define X as the number of friends a randomly selected person has.

5.5 Expected Value

The **expected value** or mean of a random variable X is the average value we expect:

$$\mathbb{E}[X] = \sum x_i \cdot P(X = x_i)$$

This gives the long-run average outcome of X.

5.6 Variance and Standard Deviation

The **variance** of X measures how spread out the values are:

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

The **standard deviation** is the square root of the variance and gives a sense of how much X typically deviates from its mean.

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

5.7 Linearity of Expectation

Even if two random variables X and Y are dependent:

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

This property is especially useful in analyzing networks.

5.8 Sampling Bias

Sampling bias occurs when some outcomes are more likely to be observed than others. In social networks, people with many connections are more likely to appear in your friend list, which affects how we interpret averages.

6 The Friendship Paradox Explained

The **Friendship Paradox** is a surprising and well-documented observation in social network theory: on average, your friends have more friends than you do.

At first, this may sound like a personal problem (it is for me), but it's actually a mathematical effect caused by the structure of social networks. The paradox arises because people with many friends are more likely to be counted in the average of "your friends" than those with few friends. This creates a bias that makes the average friend more popular than the typical person.

Let there be n people, labeled $1, 2, \dots, n$.

Let i represent a person in the population.

Let f_i be the number of friends that person i has.

1. Pick a Random Person

Define a random variable X which is a uniformly chosen person.

Then the expected number of friends that a randomly chosen person has is:

$$\mathbb{E}[f(X)] = \frac{1}{n}\sum_{i=1}^{n} f_i = \frac{f}{n}$$

This is the average number of friends in the population.

2. Pick a Random Friend

Now, imagine every person writes down the name of each of their friends on separate slips of paper. There are f total slips (one per friend-link), i.e. $f = \sum_{i=1}^{n} f_i$.

Pick one slip at random. Let Y be the person whose name is on the slip. So Y is a random friend, chosen with probability proportional to their number of appearances (i.e., f_i times).

So the probability that person i is selected is:

$$P(Y = i) = \frac{f_i}{f}$$

Then the expected number of friends that a random friend has is:

$$\mathbb{E}[f(Y)] = \sum_{i=1}^{n} f_i \cdot \frac{f_i}{f} = \frac{1}{f}\sum_{i=1}^{n} f_i^2$$

3. Comparison Using Variance

Now compare:

- Average number of friends: $\mathbb{E}[f(X)] = \frac{f}{n}$
- Average number of friends of a friend: $\mathbb{E}[f(Y)] = \frac{1}{f}\sum_{i=1}^{n} f_i^2$

Apply the variance identity:

$$\mathrm{Var}(f(X)) = \mathbb{E}[f(X)^2] - (\mathbb{E}[f(X)])^2 \geq 0 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} f_i^2 \geq \left(\frac{1}{n}\sum_{i=1}^{n} f_i\right)^2$$

Multiply both sides by $\frac{n}{f}$:

$$\frac{1}{f}\sum_{i=1}^{n} f_i^2 \geq \frac{f}{n} \Rightarrow \mathbb{E}[f(Y)] \geq \mathbb{E}[f(X)]$$

Conclusion (Friendship Paradox in Probability Terms)

A randomly selected friend is more likely to be someone with many friends, so their average number of friends is higher than that of a randomly selected person.

This inequality is strict unless everyone has the same number of friends, in which case the two expectations are equal.

7 Conditional Probability worked out example

In a small group, there are 3 people,

- Person A has 1 friend,
- Person B has 2 friends,
- Person C has 3 friends.

If you randomly select a friend from the group, what is the probability that the selected friend has 3 friends?

solution:

Step 1: List how many chances each person has to be picked as a friend:

- A can be picked 1 time,
- B can be picked 2 times,
- C can be picked 3 times.

Total friend choices = 1+2+3=6

Step 2: Find how many times a friend with 3 friends (Person C) could be picked:

- Person C has 3 chances.

Step 3: Set up the probability: P(Selected friend has 3 friends) = 3/6=1/2

8 But Why Is It Real?

Even though the paradox is mathematically sound, many people still doubt its validity. After all, how can a majority of people be below average?

The key insight is that this paradox doesn't claim that *everyone's* friends are more popular—just that, *on average*, your friends tend to have more friends than you. This is due to a form of sampling bias: people with more friends are more likely to be included in friend samples simply because they appear in more people's social circles.

In layman's terms, popular people get "overcounted" when averaging over all friends. Just like celebrities like Will Smith are more likely to be seen in public or on your feed, people with many connections dominate the social landscape, skewing the averages.

9 Applications and Implications

The friendship paradox has implications far beyond a moment of social media insecurity. It plays an important role in the following: opinion polls, disease spread modeling, social media influence, targeted marketing, detecting emerging trends, understanding peer pressure dynamics, and optimizing information dissemination in networks.

- **Epidemiology**: People who are more connected are more likely to spread diseases, so monitoring their behaviors gives early warnings for outbreaks.
- **Social Behavior**: The illusion that everyone is more popular than you can impact mental health, especially among teens.
- **Network Analysis**: Understanding biases in sampling can improve data collection in social networks and help correct misinterpretations.

10 Conclusion

The friendship paradox demonstrates how mathematical reasoning can expose surprising truths about everyday life. By understanding how averages and probability interact in networks, we reveal a bias rooted not in social failure but in the math of how connections work. This paradox serves as a powerful reminder: what feels intuitively true might just be a statistical illusion.

References/Bibliography

- Ross, S. M. (2014). *A First Course in Probability* (9th ed.). Pearson.
- Porter, Theodore M. "Probability and Statistics — History, Examples, & Facts." Encyclopædia Britannica, 2019, www.britannica.com/science/probability.

Acknowledgment

We would like to express our deepest gratitude to our amazing mentor, Kai Edick, for his exceptional mentorship throughout the course. His guidance, patience, and support have been invaluable.

Additionally, we thank Marisa Gaetz and Mary Stelow for organizing the MIT PRIMES Circle and providing us with the opportunity to deep explore probability during high school. These opportunities strengthen the next generation of mathematicians and accurately reflect MIT's commitment to fostering intellectual growth and innovation.

Thank you, MIT PRIMES Circle for cultivating an environment where curiosity thrives and inspire young mathematicians to farther their mathematical inquiry