Choose Ohio First

The Birthday Paradox

Michael Skowrons, Michelle Waugh
Dr. Artem Zvavitch

Celebrate Centennial
Kent State University
1910·2010

Introduction

We explored approaches to solving and explaining the Birthday Paradox, as well as the underlying probability theory. The problem is concerned with the probability of two people sharing a birthday within a finite group, and the solution shows this probability is surprisingly high. Solving this problem makes use of the Pigeonhole Principle, Central Limit Theorem, and numerical techniques. The solution appears counterintuitive, even paradoxical, but follows from basic probability theory.

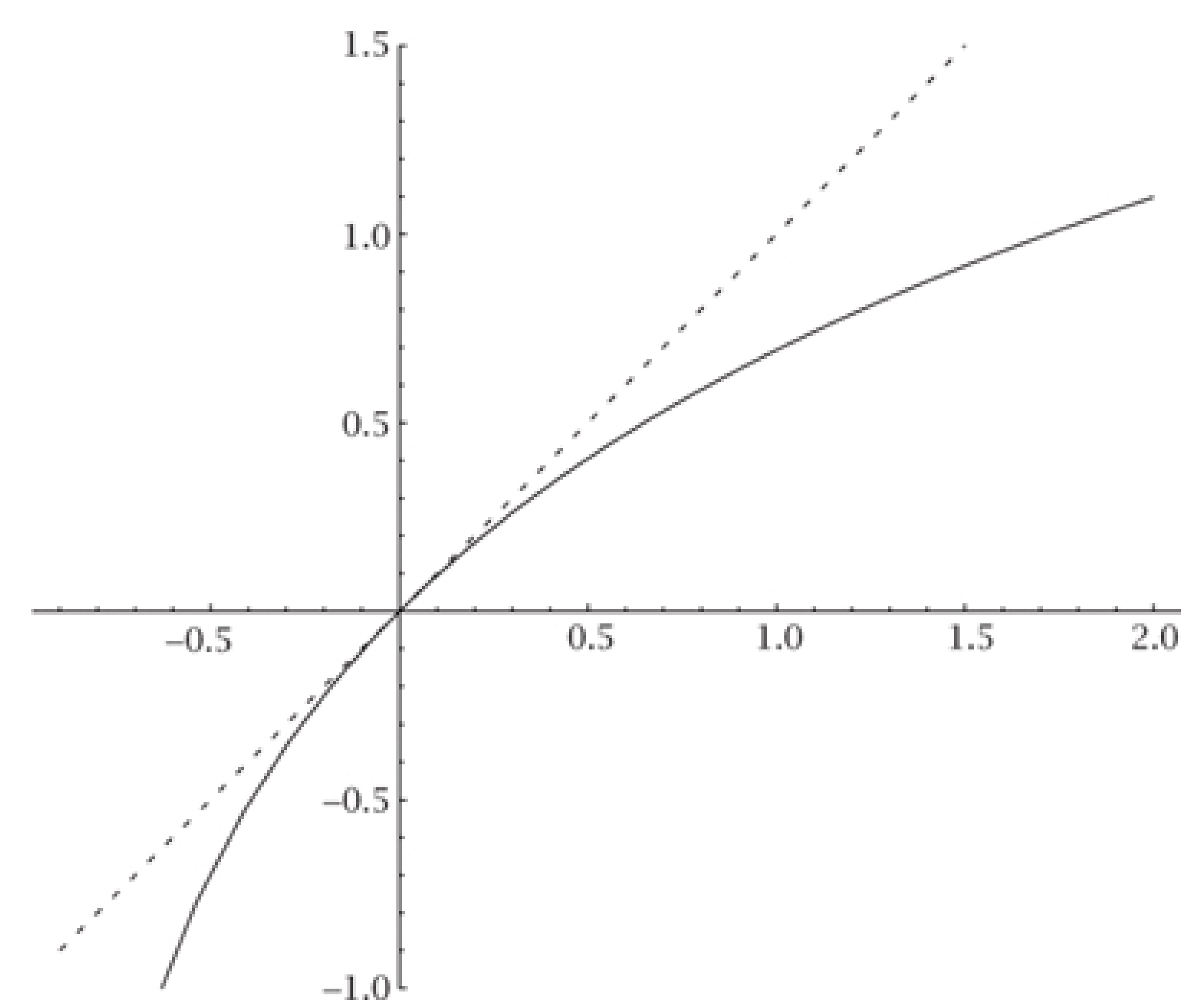
The Birthday Problem

The description of the Birthday Problem is fairly simple. Imagine there is a group of 23 people in a room. What is the chance that two of them will share a birthday? One would expect the chance to be fairly small. If there were 23 names and 365 boxes (one for each day of the year), then most of the boxes would be empty. In reality, there is a 50:50 chance that two people will share a birthday in a group. We will explain this solution, as well as the problem in general, and the underlying probability theory.

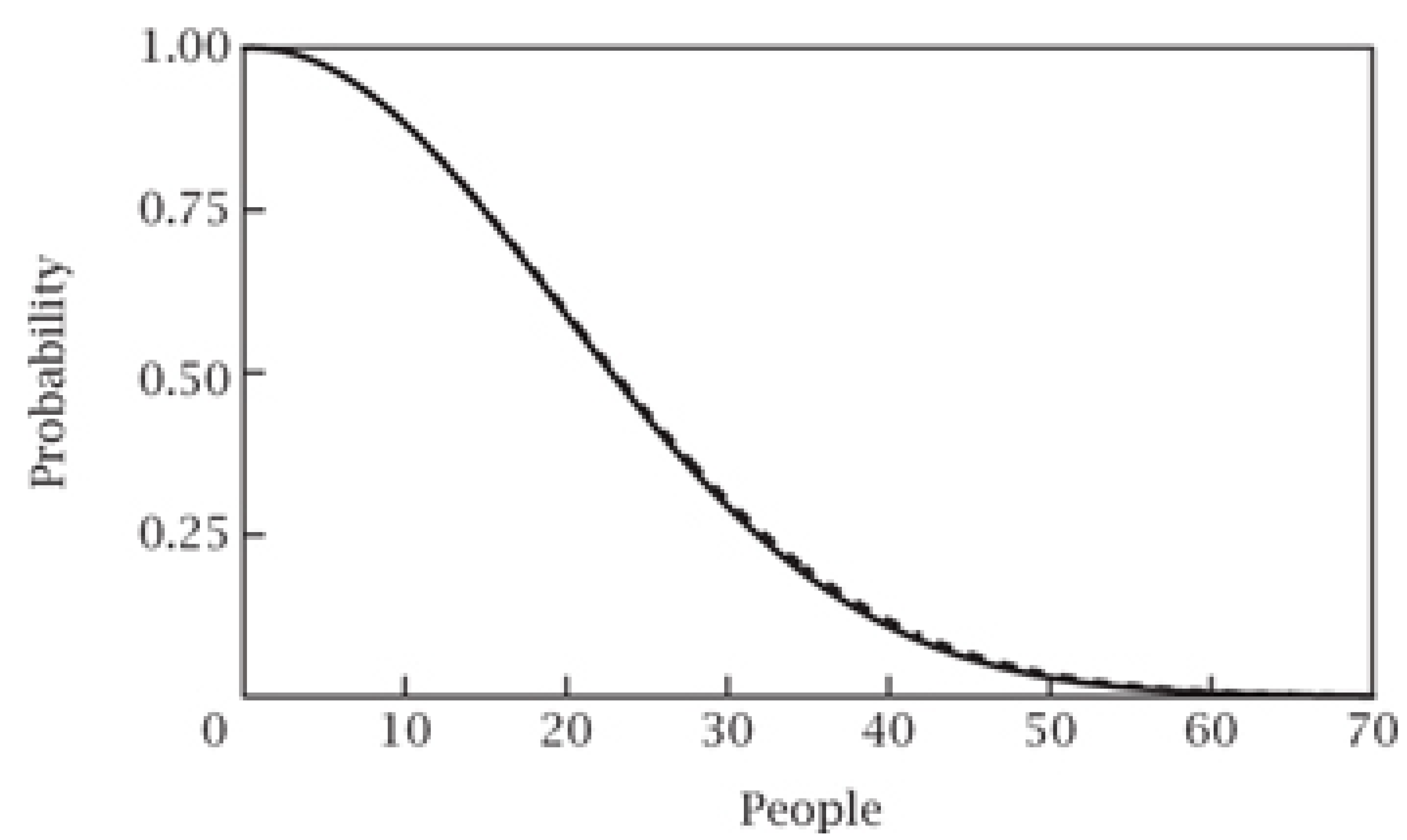
Underlying Theory

One underlying probability theory is called the pigeonhole principle. This principle basically states that if there are fewer objects than the number of boxes available for them, then at least one box must contain more than one of the objects. Introduced in 1834, it was originally created in reference to actual pigeonholes and the pigeons that went in them. Now, the principle is used to not only explain some intuitive arguments but is also used to demonstrate unexpected results. In relation to the birthday paradox, the pigeonhole principle can be used to intuitively see that as the number of people grow larger (or approach 367), at least 2 people will have to be assigned to a certain "box" (birthday) since there are only 366 possible birthdays, resulting in people having the same birthday.
Another underlying probability theory is the central limit theorem. The central limit theorem is one of the most important principles in mathematics; without it, statistical analysis would be near impossible. This theorem states that the larger the sample size from a population, the mean of all of the samples will approach the actual mean of that population which would be normally distributed. With this in mind, the samples create a bell shaped curve. The birthday paradox is related because the graph of the probability of people not having the same birthday is also normally distributed, resulting in a bell shaped curve.

Graphs



Tangent line to natural log



Probability of avoiding a match in the Birthday Problem for a set number of people. Notice the 50% chance at around 23 people, and how a match is nearly certain past around 60 people.

Solving the Paradox

Say there are k people and n dates (or boxes). The easiest way to approach this problem is to start by finding the probability of not getting a match. Imagine you are putting names of people in to boxes one-by-one. The first name would obviously have no matches. The second name would have a choice of n-1 out of n names to avoid a match, or 1 – 1/n. The third name would have a choice of n-2 out of n names to avoid a match, or 1 – 2/n. If we were to continue this trend for k people, the chance of no matches would be

$$P = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right).$$

For n much larger than k, each of these factors is close to 1. This corresponds to a low chance of matching for each factor. This formula could actually be used to solve the problem, given k and n. This doesn't say much about the trend of the probability in general though.

In order to estimate the size of the product, we can turn it into a sum by taking the log of both sides:

$$\ln P = \ln\left(1-\frac{1}{n}\right)+\ln\left(1-\frac{2}{n}\right)+\cdots+\ln\left(1-\frac{k-1}{n}\right).$$

An approximation can then be made, using that $\ln(x+1) \approx x$ for small x. This approximation can be clearly shown by looking at the graph of the natural logarithm and the tangent line about x=0. Using this approximation we obtain:

$$\ln P \approx -\frac{1}{n}-\frac{2}{n}-\cdots-\frac{k-1}{n}.$$

This is -1/n times the arithmetic progression 1+2+3+…+(k-1), which sums to:

$$\ln P \approx -\frac{k(k-1)}{2n}$$

If we then graph P, we will have a good understanding of how the probability of not finding a match varies as the number of people increases. The result is an decaying exponential trend, which would resemble a bell (normal) curve if there was a negative component on the horizontal axis.

Conclusion

The solution to this problem may seem paradoxical at first, but with an understanding of normal probability curves the answer is actually quite intuitive. Sharing a birthday in a fairly small group is actually more common than one might think. The reasoning just follows from basic probability theory.