CS 201A PROJECT

PROBABILITY PARADOXES

Project Members:
Kshiitiz Singhal(14330)
Gopichand Kotana(14249)
Gunda Abhishek(14257)
Shubam Kumar Pandey(14679)
Ishika Soni(14079)
R Animesh(14511)

Abstract

I protest against the use of infinite magnitude as something accomplished, which is never permissible in mathematics. Infinity is merely a figure of speech, the true meaning being a limit."
− C. F. Gauss
The theory of probability serves fertile grounds for puzzlement and bewilderment. An axiomatic mathematical formulation of these concepts helps us study various theory as in mathematics, statistics, finance, gambling, physics, artificial intelligence/machine learning, computer science, game theory, and philosophy to, for example, draw inferences about the expected frequency of events. Here in our project we have analyzed several paradoxes in probability.

1 Monty Hall problem

1.1 Introduction

The Monty Hall problem is a brain teaser loosely based on the American television game show, "Let's Make a Deal" and named after its original host, Monty Hall. The problem was originally posed in a letter by Steve Selvin to the American Statistician in 1975.

1.2 The Paradox

Suppose you're on a game show, and you're given a choice among three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

1.3 Solution

The answer is he should switch. This sounds so counter intuitive (1 goat and 1 car, so why switch, right?), but here I'll explain why though.

Explanation

Intuitive Explanation

An intuitive explanation is that if the contestant picks a goat (2 of 3 doors) the contestant will win the car by switching as the other goat can no longer be picked, while if the contestant picks the car (1 of 3 doors) the contestant will not win the car by switching (Carlton 2005, concluding remarks). The fact that the host subsequently reveals a goat in one of the unchosen doors changes nothing about the initial probability.From this you can see it doesn't matter whether the host reveals whats behind a door or not chances of winning by switching is always better than chances of winning by not switching.

Let's give a mathematical proof

Direct calculation

Consider the events C1, C2 and C3 indicating the car is behind respectively door 1,2 or 3. All these 3 events have probability 1/3. The player initially choosing door 1 is described by the event X1. As the first choice of the player is independent of the position of the car, also the conditional probabilities are P(Ci|X1)=1/3. The host opening door 3 is described by H3. For this event it holds:

$$P(H3|C1, X1) = 1/2$$
$$P(H3|C2, X1) = 1$$
$$P(H3|C3, X1) = 0$$

Then, if the player initially selects door 1, and the host opens door 3, the conditional probability of winning by switching is

P(C2|H3,X1)=

$$
\begin{aligned}
&\frac{P(H3|C2,X1)P(C2|X1)}{P(H3|X1)} \\
&= \frac{P(H3|C2,X1)P(C2|X1)}{P(H3|C1,X1)P(C1|X1)+P(H3|C2,X1)P(C2|X1)+P(H3|C3,X1)P(C3|X1)} \\
&= \frac{P(H3|C2,X1)}{P(H3|C1,X1)+P(H3|C2,X1)+P(H3|C3,X1)} \\
&= \frac{1}{1/2+1+0} \\
&= 2/3
\end{aligned} \tag{1}
$$

1.4 N doors

N−door generalization of the original problem in which the host opens p losing doors and then offers the player the opportunity to switch; in this variant switching wins with probability (N-1)/(N)(N-p-1). If the host opens even a single door, the player is better off switching, but, if the host opens only one door, the advantage approaches zero as N grows large.. At the other extreme, if the host opens all but one losing door the advantage increases as N grows large (the probability of winning by switching approaches 1 as N grows very large).

This give you better understanding of the fact that switching is better.What is the probability out of a 1000 doors what you had chosen initially was right and remaining were 999 were wrong . Lets take it this way that the host opens 998 of the unchosen doors and shows goats , now you get a better understanding and you would probably switch as you start doubting your initial decision or you just might know some probability .This gives you an intuition that you should switch much more clearly. I hope this gives you clear idea about Monty hall problem.
Here is something for you to think about .
What if Monty doesn't know?

2 Three prisoner's paradox

2.1 The Paradox

Three prisoner's A, B, and C have been taken in to custody for murder and their verdicts will be read and their sentences to be executed the next day. Governor has selected one of them at random to be pardoned. The warden knows which one is to be pardoned, but is not allowed to tell. Prisoner A begs the warden to tell him the name of one of the others who is going to be executed. "If B is to be pardoned, give me C's name.If C is to be pardoned, give me B's name and if I am to be pardoned flip a coin to decide whether to name B or C".
As the question is not directly about A's fate, the warden obliges and tells A that "B is to be executed". Assuming the warden's truthfulness, A is pleased that his probability of survival has gone up from 1/3 to 1/2, as it is now between him and C. A now secretly tells C about the B, who is also pleased because he reasons that A still has the probability 1/3 to be the pardoned one but his chances of survival has gone up from 1/3 to 2/3.
The problem lies with the reasoning of A about his probability of being pardoned and the facts given by C. But in reality does A gain any information from warden of his survival?

The three prisoners problem was originally mentioned by Martin Gardner in his "*Mathematical Games*" column, 1959 edition of "*Scientific American*". The three prisoner's problem surely is Monty Hall's problem in all but name with few adjustments. By simply replacing the three prisoners with the three doors, the pardon with the car, the prisoners to be executed with the doors concealing the goats and warden with Monty Hall brings back to the original Monty Hall's problem.

2.2 Solution

Let A denote the event A will be pardoned; similarly B and C denote the events of B and C being pardoned . Let I denote the event that the warden informs A that B will be executed.
Now the question becomes the probability of $\mathrm{P}(A \mid I)$?

As $\mathrm{P}(A)=\mathrm{P}(B)=\mathrm{P}(C)=1/3$ and
P(I)=P(warden saying B dies)=P(I and A being pardoned)+P(I and B being pardoned)+P(I and C being pardoned)
Mathematically,
$\mathrm{P}(I)=\mathrm{P}(I \mid A)\times\mathrm{P}(A)+\mathrm{P}(I \mid B)\times\mathrm{P}(B)+\mathrm{P}(I \mid C)\times\mathrm{P}(C)$
$\mathrm{P}(I \mid A)\times\mathrm{P}(A)=1/2\times1/3=1/6$
$\mathrm{P}(I \mid B)\times\mathrm{P}(B)=0\times1/3=0$
$\mathrm{P}(I \mid C)\times\mathrm{P}(C)=1\times1/3=1/3$
P(I)=1/3+0+1/6=1/2
By Bayes formula,
$\mathrm{P}(A \mid I)=\mathrm{P}(I \mid A)\times\mathrm{P}(A)/\mathrm{P}(I)=\frac{1/6}{1/2}=1/3$
In other words knowing B being executed, the probability of A being pardoned doesn't change and the probability of C being pardoned increased to 2/3

3 Two Envelope Problem (exchange paradox)

The problem goes like this: Given two identical envelopes one contains double the amount of money than other. With this information your friend ask you to choose one of the envelope and take a look at the money in it. Now your friend gives you a chance to switch envelopes in order to maximize your profit. Do you go for a switch or not?

Intuitively there is no point in switching because the situation is symmetric. But on calculating the expectation (which looks fair) switching seems to be a better option.

3.1 The paradox

- Suppose you find amount ₹ x in envelope.
- Now the other envelope may contain either ₹ 2x or ₹ x/2 with equal probability.
- So expected amount of money in other envelope is given by:
  E[x]=1/2*(2x)+1/2*(x/2)=₹ 5/4x.
- This suggests to us that switching is a better strategy.

3.2 Analytical Solution

There is a fallacy in argument made for expected amount calculation. Though it is true that other envelope may contain ₹ 2x or ₹ x/2.

Here it seems we are considering three envelopes but we have only two available envelops let's analyze the situation in two cases.

- Suppose we have envelope of ₹ x and ₹ 2x. Then on randomly choosing an envelope our E[x]=3/2x.
- Suppose we have envelope of ₹ x and ₹ x/2. Replace x/2 by y. Then on randomly choosing an envelope our E[y]=3/2y.
- Basically instead of having two cases we have only one and our expected income is average of amounts in envelopes always and their is no point in switching. But if you are stuck to the previous argument then their are many mathematical solution for the same.

4 Bertrand's paradox

4.1 The Paradox

For this problem let us consider a circle of unit radius. In this we inscribe an equilateral triangle. Now suppose a chord is constructed at random. The question is, what is the probability that the chord is longer than a side of the triangle? Now starts the fun part of the problem. We will give three different arguments, all apparently valid.

- **The "Endpoints" method:** Choose two random points on the circumference of the circle and draw the chord joining them. Now imagine the triangle rotated so that one of its vertex coincides with one of the chord endpoints. Observe that if the other chord endpoint lies on the arc between the endpoints of the triangle side opposite the first point, the chord is longer than a side of the triangle. The length of the arc is one third of the circumference of the circle, therefore the probability that a random chord is longer than a side of the inscribed triangle is 1/3.
- **The "Radius" method:** Choose a radius of the circle, choose a point on the radius and construct the chord through this point and perpendicular to the radius. Now rotate the triangle so that one of its sides is perpendicular to the radius. The chord is longer than a side of the triangle if the chosen point is nearer the center of the circle than the point where the side of the triangle intersects the radius. The side of the triangle bisects the radius, therefore the probability a random chord is longer than a side of the inscribed triangle is 1/2.
- **The "Midpoint" method:** Choose a point anywhere within the circle and construct a chord with the chosen point as its midpoint. The chord is longer than a side of the inscribed triangle if the chosen point falls within a concentric circle of radius 1/2 the radius of the larger circle. The area of the smaller circle is one fourth the area of the larger circle, therefore the probability a random chord is longer than a side of the inscribed triangle is 1/4.

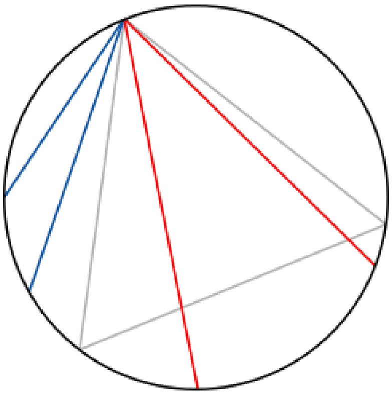

Figure 1: Random End points method Probability:1/3

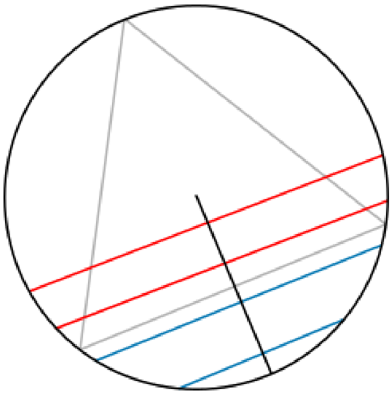

Figure 2: Random Radius method Probability: 1/2

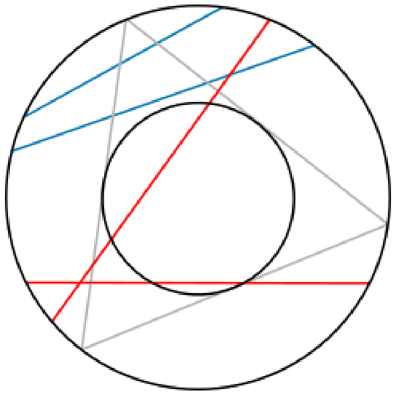

Figure 3: Random Midpoint method Probability: 1/4

You can try all you want to, but you wont find any mathematical inconsistencies in the three methods described above. So where does our problem arise??

4.2 The Solution

- **The Classical Solution:** Our solution hinges on what we consider random, which in this case boils down to, "Which method truly gives us random distributions?". So before we proceed any further let us introduce the "Principle of Indifference" given by Laplace. Suppose that there are n>1 mutually exclusive and collectively exhaustive possibilities. The Principle of Indifference states that if the n possibilities are indistinguishable except for their names, then each possibility should be assigned a probability equal to 1/n.

  For example: A regular symmetric dice has 6 faces, arbitrarily labeled from 1 to 6. We assume that the die must land on one face or another (Mutually Exclusive), and there are no other possible outcomes (Exhaustive). Applying the principle of indifference, we assign each of the possible outcomes a probability of 1/6.

  In our analysis of Bertrand's Paradox, we misuse this principle. The nonsensical results are due to multivariate, continuous variables. While our three probability distribution seems to be uniform in their individual situations cases but when viewed for the other properties turn non uniform.

  For instance, each of the three selection methods presented above yields a different distribution of midpoints. Methods 1 and 2 yield two different non uniform distributions, while method 3 yields a uniform distribution. On the other hand, if one looks at the images of the chords below, the chords of method 2 give the circle a homogeneously shaded look, while method 1 and 3 do not.

- **Jaynes' solution using the "maximum ignorance" principle:** In 1973, Edward Jaynes proposed that a solution to the Bertrand's Paradox must be indifferent to scale and translation invariant. That is if I construct any circle of half the size inside our unit circle and keeping the chord distribution to be the same, our answer should not change. Considering his proposition we see that the chord distribution

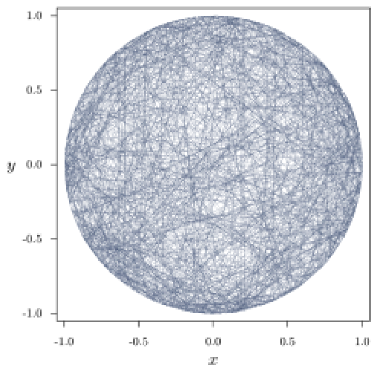
Figure 4: Method 1

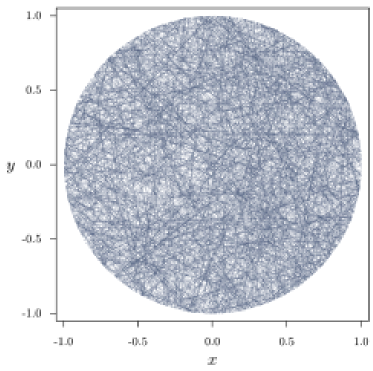
Figure 5: Method 2

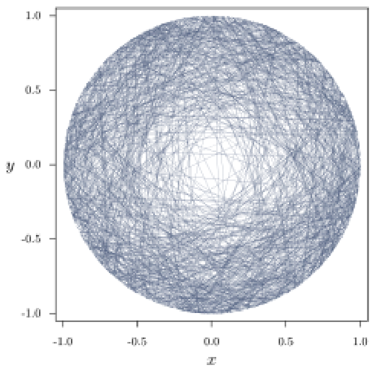
Figure 6: Method 3

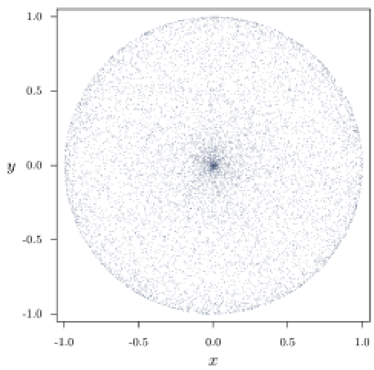
Figure 7: Method 1

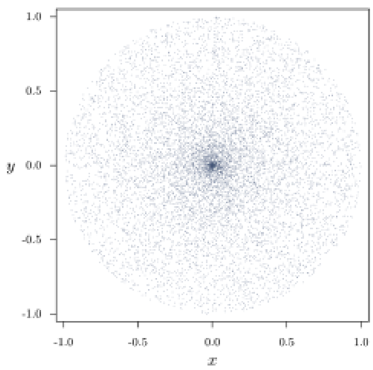
Figure 8: Method 2

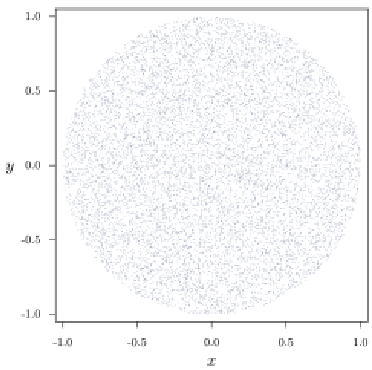
Figure 9: Method 3

changes in the smaller circle if we pick the chords by method 1 or method 3. Hence according to Jaynes only method 2 gives us the true solution to our problem.

5 Simpson's Paradox

It's a well accepted fact that the larger the data set, the more reliable the conclusions drawn. Simpson' paradox, however, contradicts and the result is a good deal worse than a sore thumb. Simpson's paradox demonstrates that a great deal of care has to be taken when combining small data sets into a large one. Sometimes conclusions from the large data set are exactly the opposite of conclusion from the smaller sets. Unfortunately, the conclusions from the large set are also usually wrong.
The problem can be illustrated through the following example:
In 1973, the University of California-Berkeley was sued for gender discrimination against women who applied for admission to undergraduate schools there.

Combining different data-sets, sometimes results in reversal of disappearing of existing and prevailing data trends.
There are many interesting counter-intuitive results we get to notice once we know the Simpson's paradox, let's see them through some examples;

Table 1: Admission Figures

	Applicants	Admitted %
Men	8442	44%
Women	4321	35%

The given figures in the above table clearly showed the large difference in the percentage of men and women who got into the university. This, by no chance seemed a coincidence.
But when we examined the data from individual departments, as in the table below, of the university, the results were difficult to believe. It was observed that the no department was significantly biased against women but to everyone's surprise a majority of departments were significantly biased in favor of women.

Table 1: Data From Six Largest Departments of 1973 Berkeley Discrimination Case

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

Source: Bickel, Hammel, and O'Connell (1975); table accessed via Wikipedia at https://en.wikipedia.org/wiki/Simpson%27s_paradox.

There are two drugs, drug A and drug B which cure the same disease,they have been tested on two different days and these are the statistics obtained:

Table 2: Number of people cured by Drug A and Drug B on day one and day two respectively

Drug-A	Drug-B
25/30 (85.33%)	60/90(66.67%)
40/120 (33.33%)	10/60(16.67%)

The total number of people that they cure is as following,Drug$A$ 65(43.33%) and Drug$B$ 70(46.67%). When a doctor is presented with these statistics he should go with the Drug$B$ even though Drug$A$ has more percentage of people cured on both the days this is because the lurking variable is unknown in this case, there might be so many factors playing a role in the drhttps://preview.overleaf.com/public/cqfqzhytznhk/images/a1cdf70708b53b816d19351a6f4cde9f011 ability to cure a person,so it's a better decision to go with the overall percentage. To give a clear emphasize we present you with a new table (Table 3)

Clearly, Drug$B$ is way more preferable than Drug$A$

Now let's say they are very common drugs and the scientists were able to figure out the lurking variable, let's say these drugs behave exactly in this, that is cure the same percentage of people irrespective of the number of

Table 3: Number of people cured by Drug A and Drug B on day one and day two respectively

Drug-A	Drug-B
1/1(100%)	98/99(98.9%)
1/99(1.01%)	0/1(0.00%)

people it is tested on, then clearly drug$A$ is a better drug to use as any day it will cure more more percentage of people irrespective of the number of people to whom the drug is given. You are a marketing executive and it's profitable to sell Drug B then you test the drugs on the number of people as mentioned in the tables so that it looks as if the drug B cures more number of people in total and doesn't disclose the fact that drug A will perform better any day. In this way you can manipulate data to cater your needs.

Simpson's Paradox is an effect of a combination of a lurking variable and data from unequal sized groups being combined into a single data set. The unequal group sizes, in the presence of a lurking variable, can weight the results incorrectly. This can lead to seriously flawed conclusions. In order to prevent such misleading results we should not combine data sets of different sizes from a diverse sources.

Simpson's Paradox will generally not be a problem in a well designed experiment or survey if possible lurking variables are identified ahead of time and properly controlled. This includes eliminating them, holding them constant for all groups or making them part of the study.

6 Non Transitive Dices

Transitivity : In terms of set theory, the transitive relation can be defined as: Given a set X for all a,b,c in X (aRb & bRc) implies aRc

A non-transitive game is something in which if A wins over B, B wins over C then it does not imply A wins over C. Non- transitive games are common, and we have played some of them like ROCK-PAPER-SCISSORS.

The non-transitive dice is particularly interesting because, the non-transitiveness is not absolute but probabilistic, that is, say we have defined a win as getting a number more than the other dice then, probability of Dice A winning is more than B winning, probability of B winning is more than C winning then probability of A winning need not be more than C winning.It is possible to find sets of dice with the even stronger property, that is probability of C winning is more than A winning. So, if you play the game many times you would expect A to win over B, B to win over C and C to win over A. Using such a set of dice, one can invent games which are biased in ways that people unused to non-transitive dice might not expect

Let's device some games now:

- Die A has sides 2,2,4,4,9,9.
- Die B has sides 1,1,6,6,8,8.
- Die C has sides 3,3,5,5,7,7.

Rule of the game : The dice which has a higher number appearing on it wins.z

Let's paint the surfaces of the dice differently so that we can differentiate between the sides which has the same numbers. After doing this let's roll two dice at a time, there are 36 cases which are possible, any one of the 6 sides on Dice 1 and any one of 6 sides on Dice2 let's calculate the probability of:
A winning over B, the possible cases where A can win over B are

[(2,1),(2,1),(2,1),(2,1),(4,1),(4,1),(4,1),(4,1),(9,1),(9,1),(9,1),(9,1),(9,6),(9,6),(9,6),(9,6),(9,8),(9,8),(9,8),(9,8)] Total 20 cases are possible therefore the probability 20/36=5/9 B winning over C, the possible cases where B can win over C are [(6,3),(6,3),(6,3),(6,3),(6,5),(6,5),(6,5),(6,5),(8,3),(8,3),(8,3),(8,3),(8,5),(8,5),(8,5),(8,5),(8,7),(8,7),(8,7),(8,7)] Total 20 cases are possible therefore the probability 20/36=5/9

C winning over A, the possible cases where C can win over A are

[(3,2),(3,2),(3,2),(3,2),(5,2),(5,2),(5,2),(5,2),(5,4),(5,4),(5,4),(5,4),(7,2),(7,2),(7,2),(7,2),(7,4),(7,4),(7,4),(7,4)] Total 20 cases are possible therefore the probability 20/36=5/9

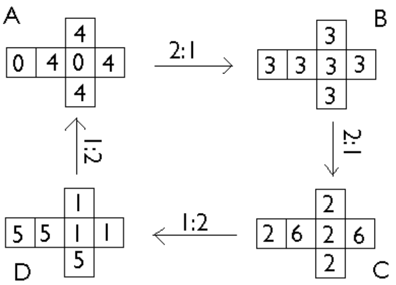

Figure 10: Four Non Transitive Dices

Example: Grime dice

- dice A has sides 2,2,2,7,7,7
- dice B has sides 1,1,6,6,6,6
- dice C has sides 0,5,5,5,5,5
- dice D has sides 4,4,4,4,4,9
- dice E has sides 3,3,3,3,8,8

A beats B beats C beats D beats E beats A (first chain);
A beats C beats E beats B beats D beats A (second chain).