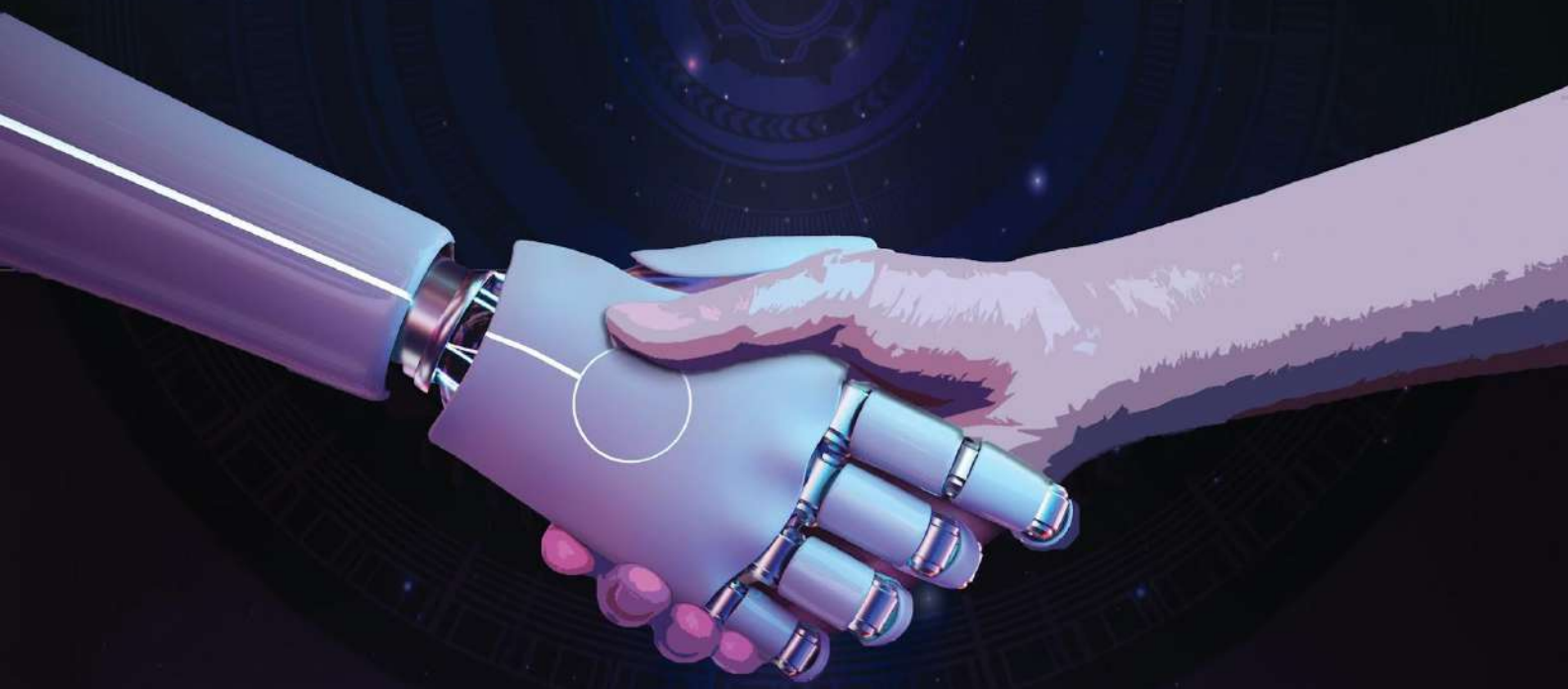
St. Xavier's College (Autonomous), Kolkata

DEPARTMENT OF STATISTICS



PRAKARSHO

VOLUME XVII

2025

St. Xavier's College (Autonomous), Kolkata

DEPARTMENT OF STATISTICS

PRAKARSHO

VOLUME XVII

Scan to get e-copy

2025

Department of Statistics

Left to right: Prof. Rahul Roy, Dr. Surupa Chakraborty, Dr. Surabhi Dasgupta, Dr. Durba Bhattacharya, Prof. Madhura Das Gupta, Dr. Soumya Banerjee, Dr. Sancharee Basak,

Undergraduate Department of Statistics

Batch of 2022-2025

Batch of 2023-2027

Batch of 2024-2028

Postgraduate Department of Data Science

Batch of 2023-2025

Batch of 2024-2026

Contents

Message from the Principal

Rev. Dr. Dominic Savio, SJ
Principal
St. Xavier's College (Autonomous), Kolkata

*"*I'm delighted to announce that the Department of Statistics of our esteemed college is gearing up to unveil the highly anticipated 17th edition of its annual magazine 'PRAKARSHO.'

The departments' steadfast dedication to excellence radiates through the pages of this magazine. Since its establishment, the Department of Statistics has served as a vibrant hub, nurturing innovative skills and fostering a fervor for research. With the addition of the Postgraduate Department of Data Science, the development of research and innovation has only reached a higher level of excellence. Through this magazine, students from diverse fields are encouraged to contribute articles pertaining to statistics and data science, fostering a culture of exploration beyond their academic disciplines.

Congratulations to the department—both the esteemed faculty and the exceptional students! Here's to the triumphant release of this edition and the assurance of continued inspiration for future endeavors. Wishing boundless success to everyone involved!

Nihil Ultra!"

Principal

Message from the

Vice Principal

Prof. Bertram Da Silva
Vice Principal, Arts & Science
St. Xavier's College (Autonomous), Kolkata

"For decades, the Department of Statistics at St. Xavier's College (Autonomous), Kolkata, has been a beacon of research and innovation. The incorporation of the Postgraduate Department of Data Science has enhanced the scope of knowledge sharing and intellectual growth within the department. This enduring commitment to cultivating a culture of learning shines brightly through the pages of the 17th edition of their departmental magazine, 'PRAKARSHO'.

The magazine is a platform for the students and researchers from diverse corners of the country converge, engaging thought-provoking discussions and sharing ground-breaking statistical theories. A rich tapestry of articles, research papers and captivating segments grace its pages, inspiring students to dive deep into the ever- evolving realm of science.

I congratulate the students and faculty members for the successful launch of the magazine."

Vice Principal

Message from the

Dean of Science

Dr. Indranath Chaudhuri
Dean of Science
St. Xavier's College (Autonomous), Kolkata

*"*It's truly uplifting to hear about the imminent release of the 17th edition of 'PRAKARSHO', the annual magazine of the Department of Statistics. The articles in the magazine eloquently portray students' passion, drive, and commitment to their field, showcasing their exceptional dedication and brilliance. Indeed, it serves as a window into the minds of the country's budding statisticians and future data scientists, revealing their innovative thinking and aspirations.

I extend my warmest commendations to the team for the success of this endeavor. I eagerly anticipate their continued achievements and the positive impact they will undoubtedly bring to the field in future."

Dean of Science

Message from the

Head, Department of Statistics

Dr. Durba Bhattacharya
Head, Department of Statistics
St. Xavier's College (Autonomous), Kolkata

"It fills me with immense pride and satisfaction to witness the students of the Department of Statistics and the Postgraduate Department of Data Science work together in successfully presenting the 17th edition of our Departmental Magazine, PRAKARSHO. Once again, we've captured the essence of our department within its pages, a feat made possible by the unwavering dedication and relentless efforts of our students. I extend my heartfelt gratitude to Father Principal, Vice Principal, Dean of Science, and Dean of Arts for their ongoing guidance and encouragement. Special thanks are due to the Programme and Publication Committee for their invaluable support.

I commend the Student Editorial Board and the Publication Committee for their exceptional diligence, enthusiasm, and commitment in overcoming all obstacles to bring this issue to fruition. I also express my sincere appreciation to my colleagues whose collective dedication and teamwork have contributed significantly to this achievement."

Head of the Department

Message from the

Editor's Desk

"Welcome to the latest issue of the annual magazine of the Department of Statistics, St. Xavier`s College (Autonomous) Kolkata, where we delve into the intriguing world of defining abstraction. As statisticians, we are constantly facing the challenge of extraction of meaningful pattern from data, operating in realms where abstraction reigns supreme.

From deciphering the hidden structures within datasets to extrapolating trends that transcend the tangible, our contributors have illuminated the diverse ways in which statistical methods manifest in the world.

We hope that this issue sparks new ideas, fosters lively discussion, and inspires you to push the boundaries of inference in your own research and practice. Thank you for joining us on this intellectual journey. May the insight within the pages of this magazine enhance further exploration in the statistical arena.

With guidance and support from our professors, we present to you PRAKARSHO VOLUME XVII."

Nihil Ultra!"

Snija Upadhyay.

Associate Editor-in-Chief

Anisha Sengupta

Editor-in-Chief

Editorial Board

Patron

Rev. Dr. Dominic Savio, SJ
Principal

Advisory Board

Prof. Bertram Da' Silva
Vice-Principal, Arts & Science

Dr. Indranath Chaudhuri
Dean of Science

Dr.Farhat Bano
Dean of Arts

Dr. Surabhi Dasgupta
Dr. Surupa Chakraborty
Prof. Madhura Das Gupta
Prof. Rahul Roy

Dr. Ayan Chandra
Dr. Durba Bhattacharya
Dr. Sancharee Basak
Dr. Soumya Banerjee

Amisha Sengupta
Editor-in-Chief

Srija Upadhyay
Associate Editor-in-Chief

Editorial and Designing Board

EDITORIAL

Pragnya Bhattacharya
Shreyosee Sen
Shreshtha Sengupta
Sumeet Sikdar
Rishika Ghosh
Ankita Debnath
Soumyojit Das
Soumyadipto Das

Madhusree Bhattacharjee
Peyasi Mondal
Ankita Sarkar
Rasika Agarwala
Aishi Dey
Akash Roy
Anubhav Hazra

DESIGN

Sreyasi Dey
Rasika Agarwala
Arnabi Sarkar
Tamaghno Saha
Jane Vandita Toppo

Sumeet Sikdar
Daniel Dibyajyoti Mondal
Shreetama Dey
Shristi Dadel

Postgraduate Department of Data Science

Departmental Activities

INVITED TALK

(i) Biswajit Pal, Director of Data Science at Kenvue, delivered a session on 'Crafting a Standout Data Science Portfolio' on 1st March, 2025.

(ii) Mr. Sanjoy Chatterjee, Chairperson, NASSCOM Regional Council, East and Co-Founder, Entiovi Technologies delivered an invited talk on "The Future of Data Science: Industry Trends, Technological Innovations, and Emerging Careers" on 25th February, 2025.

(iii) Dr. Trambak Banerjee, Assistant Professor, School of Business, University of Kansas, delivered a session regarding PhD in the US on 7th November,2024.

(iv) Mr. Rahul Bhattacharya, AI Leader at EY GDS delivered an invited talk on 'Artificial Intelligence for a Better Working World on 2nd May,2024.

Departmental Activities

SPECIAL LECTURE SERIES

A Special lecture Series by Prof. Mausumi Bose,Visiting Professor from the National Board of Higher Mathematics (NBHM), was organized throughout the year for the Postgraduate Department of Data Science and Postgraduate and Research Department of Microbiology.

Classes are delivered by the experts from the industry: Saurav Ghosh,Vice President Data Science, Accenture Al, for MSc. Data Science students.

Multiple discussion sessions were held throughout the year with Mr. Srijit Mondal, the Team Lead –Financial Market Analysis –USA-Future's First Former Senior, former Data Scientist at Sony Liv and Zomato.

Departmental Activities

ICDMAI 2025

Students from the Postgraduate Department of Data Science and the Postgraduate and Research Department of Computer Science actively participated in the 9th International Conference on Data Management, Analytics, and Innovation, organized by the Society for Data Science in collaboration with St. Xavier's College (Autonomous), Kolkata.

Departmental Activities

EPSILON DELTA 2024
Inferring Abstraction

The Department conducted its Annual program "Epsilon Delta" on April 9th ,2024. The program commenced with the launch of the **16th edition of the Departmental Magazine Prakarsho.**

The following programs were organized through the day:

(i) Paper presentation event by students - **Proectura**
(ii) Science quiz - **Inquizzitive**
(iii) Data visualization event - **Analyticon**
(iv) Hackathon event based on a statistical theme – **Datathon**
(v) Chess - **Checkmate**
(vi) Cultural program by the students of the Department.

Undergraduate Department of Statistics

Departmental Activities

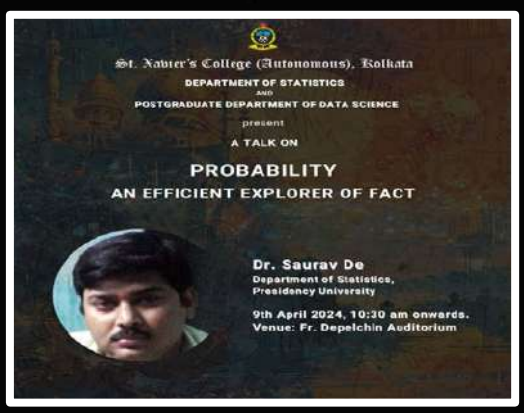
INVITED TALK

(i) Prof. Nitis Mukhopadhyay, Department of Statistics, University of Connecticut-Storrs,USA delivered a session on 'An example of Fleecing of Funds: Could Statistics Help? ' on 10th January, 2025.

(ii) Dr. Soumik Purkayastha spoke on "Journey from Xavier's to ISI to Pittsburgh, career and PhD related." on 19th November, 2024.

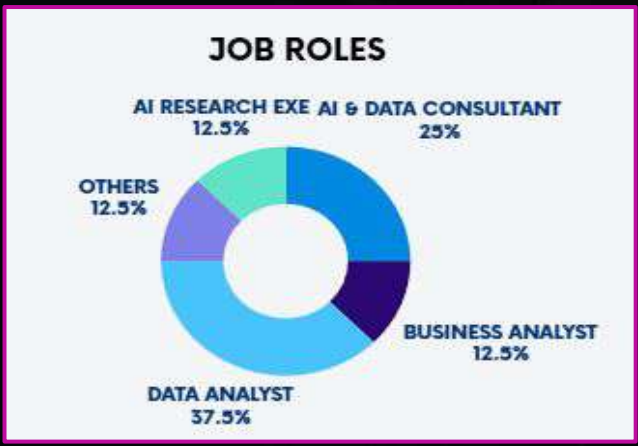
(iii) Prof. Bimal Sinha, Department of Mathematics and Statistics, University of Maryland, Baltimore County held an online seminar on "Confidence Ellipsoids of a Multivariate Normal Mean Vector Based on Noise Perturbed and Synthetic Data with Applications" on 18th August and 21st September, 2024.

(iv) Dr. Sourav De from the Department of Statistics, Presidency University delivered an invited talk on "Probability -An Efficient Explorer of Fact" on 9th April, 2024.



PLACEMENT REPORT

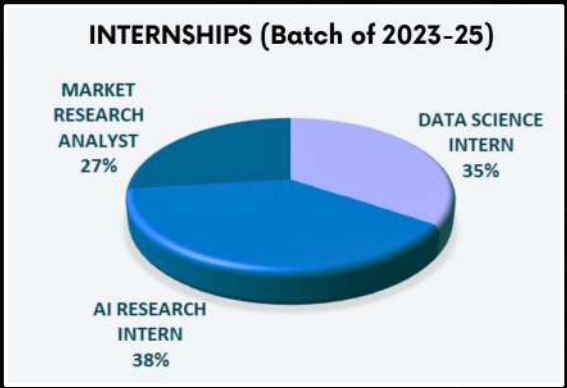
The placement drive is progressing steadily, with active participation from esteemed companies, showcasing the strong demand and recognition of our programs.



Department	Name of the Student	Organization Placed in
MSc. Data Science	Silvia Mallick	Accenture
MSc. Data Science	Shruti Jana	Accenture
MSc. Data Science	Amisha Sengupta	RSA
MSc. Data Science	Pragnya Bhattacharya	Feedsense
MSc. Data Science	Preetanwita Sarkar	EY GDS
MSc. Data Science	Sohom Ghosh	EY GDS
MSc. Data Science	Rishav Kumar	Prime Infoserv
MSc. Data Science	Vishal Acharya	Rinex Technologies
BSc. Statistics Hons.	Vishesh Modi	Accenture

INTERNSHIP REPORTS

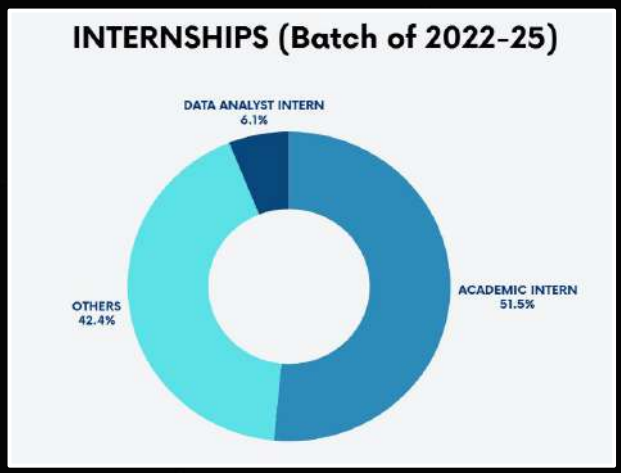
Postgraduate Department of Data Science



Approximately 80% of the students bagged internship opportunities. The postgraduate data science students exhibited a clear shift toward technology-driven roles. **AI Research Internships (38%)** were the most popular, followed by **Data Science Internships (35%)** and **Market Research Analyst positions (27%)**. This distribution showcases the growing importance of artificial intelligence, machine learning, and data-driven decision-making in modern industries.

Department of Statistics

For the undergraduate students, 76% of the students have got internship opportunities in which **Academic Internships** dominated, accounting for **51.5%**, followed by **Other internships (42.4%)**, and a smaller proportion opting for **Data Analyst roles (6.1%)**. This indicates a strong inclination towards research-oriented opportunities and statistical applications in diverse fields.



Student's Achievements

MSc. Data Science

- **SIT ICOE Hackathon**

Team CamCoderr comprising of Karan Mehta, Priti Jhalaria, Madhusree Bhattacharjee, Akash Kumar and Archisman Rakshit from the **Postgraduate department of Data Science** secured a position among the top 10 teams out of 400+ in the **SIT ICOE Hackathon**.

They grabbed the opportunity to participate in an invite-only SAP Inside Track Kolkata Community Event, an exclusive gathering of SAP enthusiasts, professionals and innovators which was held at Biswa Bangla Convention Centre on 27th April, 2024.

The event and certification partners were the International Center of Excellence for Data Science, AI, and Futuristic Technologies, in collaboration with the SIT Community, Kolkata; Alien Brains Educations; Westernacher Consulting and QBAdvisory.

BSc. Statistics (Hons.)

- **ISI MStat. 2024**

Students from the Undergraduate Department of Statistics demonstrated outstanding performance in ISI MStat. 2024 Examination

Name of the Student	All India Rank
Bishwayan Ghosh	**2**
Dyuti Manik	**10**
Dishari Datta	**21**
Baidurya Bhattacharya	**28**

Student's Achievements

- **IIT JAM 2024**

Students from the Undergraduate Department of Statistics demonstrated outstanding performance in IIT JAM 2024.

Name of the Student	All India Rank
Bishwayan Ghosh	**1**
Arka Roy	**23**
Dyuti Manik	**44**
Ayan Saha	**50**
Dishari Datta	**70**
Baidurya Bhattacharya	**77**

- **IIT JAM 2025**

Students from the Undergraduate Department of Statistics demonstrated outstanding performance in IIT JAM 2025.

Name of the Student	All India Rank
Shrestha Mukherjee	**4**
Soumyojit Das	**9**
Akash Roy	**25**
Arshiya Paul	**40**
Sreyasi Dey	**51**
Rohan Mondal	**55**
Sumedha Banerjee	**57**
Anubhav Hazra	**61**

In Conversation with…..

Dr. Amritaputra Bhattacharyya

Sr. Data and Applied Scientist at Microsoft

Dr. Amritaputra Bhattacharyya, is *Sr. Data and Applied Scientist at Microsoft.* Eager to tap into his extensive experience, Pratiksha Banisitti and Pratyusha Mukherjee, students from the Statistics and Data Science Departments, interviewed him via Microsoft Teams. This interview was conducted with the backing of the Editorial Committee.
The following is the written and abridged form of the interview , that has been prepared, in consultation with Dr. Amritaputra Bhattacharyya.

Q1. Sir, can you share some memories from your childhood and early schooling days?

Sir: Yeah, I grew up in Kolkata. So, my entire childhood, hearing about St. Xavier's, Lamarts and all the schools and colleges around my place. I studied in a school, which was also on Park Street. And I did my secondary Madhyamik from that school. I was part of the Boy Scouts in that school. Being a part of a Boy Scout, we used to meet or come to Xavier's for different purposes. Like, we used to have key events, and the Boy Scout events or a football game. We used to participate. So, it was always a fascination – "Hey, this is such a nice school!" It has two big grounds for football and stuff. It was very memorable since childhood days that, St. Xavier's is one of those colleges that you would like to join later in life and that is how it all came. Since my childhood days, I was a little bit more, probably more comfortable in physics and math and that was the main subject I thought that probably I am going to pursue at a later part in my life.

Q2. What motivated you to choose Statistics as your undergraduate major?

Sir: Yeah, so there's a Bengali magazine, weekly, I'm sure it is still there. It's called Desh. When I was in middle school, I had a chance to read about something in Statistics in Bengali called Rashibigyan. So, I had a chance to read an article about application of Rashibigyan or Statistics in different fields, how it is used from astronomy to physics and chemistry, how people are invading computer science, and how people are using it. The main thing that was fascinating for me, was that given the data, you can do some kind of prediction. In my childhood days, it used to be something like astrology. I mean, you can predict something just by seeing the data.Because at that point of time in life, astrology was like, oh, people can predict your future! You could do magic, just by seeing the data, that what can happen with some kind of relative confidence and stuff. Then I told my grandfather that I wanted to study Statistics. But that was back in school, and you never know after 11th or 12th, where life would take you. Also, through that article, I came to know about PC Mahalanobis and how Indian Statistical Institute (ISI) works. There was never any pressure from my parents that I had to be a doctor or an engineer. So, at that time I realized, and I decided that, okay, I have to be in ISI. That is the space I must go. And that is the main reason.

Dr. Amritaputra Bhattacharyya

Sr. Data and Applied Scientist at Microsoft

Q3. Can you share some memories as a student of the first batch of Statistics(Hons.) at St. Xavier's College?

Sir: Yeah. So, before I joined Xavier's, that year I had already got admission in Mathematics honours. And actually, I did my 11th and 12th from a college called Maulana Azad College. In 11th and 12th, I had Statistics as my fourth subject other than Physics, Chemistry and Mathematics. My teachers in that department at Maulana Azad College knew about my interest in that subject. So, I was very disheartened when I didn't get into Presidency and Ashutosh College because those two colleges were the only ones that used to offer Statistics at that time. And BStat, I couldn't make it. Then I heard that Xavier's is coming up with a statistics department. And I think with Xavier's there was another college, probably Bethune, I'm not sure, maybe. And so, I reached out to my professor at Maulana Azad, one of the professors over there. And I know two of them. Their initials were BR and PD. They wrote a recommendation letter to Professor Amit Ghosh. I took the recommendation letter, and I still remember I came to Xavier's and met him. Those days, there was no examination, no test for me to get enrolled. It was all based-on marks. I think they had already started classes nine or ten days before and I was on the waiting list. So, my professors at Maulana Azad sent me a recommendation letter. I went and met Dr. Amit Ghosh, and he looked at the letter and looking at my waiting list rank he said, "Okay, come tomorrow." The next day, I went directly to the Dean's office and everything was done. And I joined my class. I was grateful for those two professors who taught me statistics in Maulana Azad. BR and PD and definitely to Dr. Amit Ghosh. So, that is how it all started. And coming from a complete boys' school to Xavier's, it was like people here at that point of time, Xavier's Bombay, Xavier's Calcutta, they were quite advanced culture-wise. So, it was a little bit different, though I grew up in Beck Bagan and Park Circus, as I was telling you. But I always had good memories from that college, like clean benches. One thing I also realized was that the first kind of normalization that happened to me was that you could be the star in your school, but there are stars like you from other schools as well. So now face them. So, it's such an elite school and people are coming from all over the place and they're equally good. And I realized that, okay, now I'm not the only one. There are people like me, who are much better than me and they're also here. And that helped me to keep my feet on the ground. And I still love and cherish some of the friends I still have from those days. They're at good places, well-established, coming from a different background, but still, we are good friends. That was the first normalization that happened in my career.

Q4. What are some of the most valuable lessons or skills you've learned in your career?

Sir: I would suggest that keep a plan and don't take the smaller failures in life so seriously, failures will come and believe me I was never a bright student. I had other friends who were much better than me, who still are definitely. And, I mean, if you see after

Dr. Amritaputra Bhattacharyya

Sr. Data and Applied Scientist at Microsoft

St. Xavier's, I had some family issues and I had to get a job and I was thinking about joining some kind of a professional course so that I could start my earning fast, mainly. So, I sat for different exams, I was sitting for CDAS and even CDS, I did MStat again but couldn't make it. Then somehow, I made it to IIT Bombay. After IIT Bombay, I joined as a software engineer. But I knew that this was not the job that I would be doing for the rest of my life, and I just took it up because I had to get some income from that. So, I joined a software company but at the same time, I always had in mind that I was going to go to a domain where I could use my education in Statistics. So, then I gradually went to the company called SaaS, the software company that made the language SaaS. I knew that once I was in SaaS then I could probably do something in that domain, so I tried to join that. Then I thought okay to do something more, would you like to go to the US, and I said OK let's do a Master's. I did another Master's in the US and then again, I joined SaaS, but this time I joined the team where they were using statistics or data science to work. See, it took me some time but eventually I aligned myself with the dream I had. I wanted to work in this domain. So, what I would suggest to all the kids, there could be an initial or intermittent failure, so don't be in so much bother, in the long run, things will normalize as long as you have a plan for what you want to be or want to do in your future, it is more about that, just working towards that every day a little bit is enough work, yeah.

Q5. What advice would you give to a student of Statistics regarding the choice of his future career?

Sir: I think you are taught some kind of a programming language, during our time it was not there, and believe me coming from 11th,12th and undergrads even from Xavier's when we went to different other institutes, the national institutes. We were shocked to see that people from our age group knew so much more about computers than us and believe me, when I was given access to the server, where we were supposed to work, I didn't know what to do. I was in front of a black screen with some prompts and I didn't know what to do with that prompt. I had to learn those unique commands to manoeuvre around the folders, I was taught or I read or I somehow managed to understand you can open some files, this is a way you open a file, I mean from changing the permission and using the files for writing programmes and how to run, it was all coming as big huge shift in my career. I used to go to a friend's place, and we used to play F1 car racing and for me those up and down, right, and left buttons were the only four keys known to me. I did not do any kind of writing as well. So, one thing is, I have heard that you guys have some kind of a programming language that will give you a better place in your career, other than that given the recent development that is happening if you are interested in pursuing as a professional or in the education world, based on education, I would suggest that since I am also not from that domain, I mean, I didn't pursue my PhD but whatever I understood is that you have to be a little bit more theoretical person. You should know why things are happening, mainly the Math part of those deductions. I think those are very important and, it would be better suited for the kids or the people who are coming from the Statistics background. There's no hard and fast rule. You can also join the professional world as well, in terms of this, any big tech companies or

those places as well. But the thing is that data science is more of an applied field so there are PhD courses in the US but there are not as many as if you want to do a PhD in Statistics. So, Data Science is mainly catering towards your career by joining some kind of a job. So, on the data science side, I would suggest that try to understand the maths and how to code. You don't have to write the code from scratch. Given that the advancement that we are recently observing, there are applications called Copilot, and you have heard about ChatGPT and stuff, they are going to help you a lot, but you should know, even ChatGPT makes mistakes. So, don't take whatever it is suggesting as the truth, maybe it is good in language but may not be good in other things. So, for data science work I would suggest you understand what is going behind. How someone develops that, be it a model based on what kind of data they have built the model or the architecture of the model. Those things are very important. I am not asking you to go and understand the ChatGPT framework. I am just saying that even if you understand that normal deep learning framework, how it works, and how it has evolved, it works, and how it has evolved, that will help. And by no means I am not saying that don't do Statistics. Data Science people always say that Data Science is good, but when it comes to Linear Algebra and Statistics, people get scared so you guys are the better players and you will understand the linear algebra behind it - the optimization techniques going on behind it, why you are doing the optimization technique, why MLP, why not OLS or map estimation. So, that knowledge will help you to move forward or excel forward compared to others who are just calling APIs to fit models for the sake of data and try to be in that domain, that is where you will be at the edge.

Q6. Can you share a brief profile of your career?

Sir: When I joined the workforce in 2001, it was the big tech bubble burst, all the dot coms were failing and at that time other than GE CAPITAL, even GENPACT came along much later, there were not many companies who were giving that much of thought towards the data. There were companies like Godrej, which have sales data, right, but they used to take the tribal knowledge to incorporate a typing point or forecasting different techniques used to happen to those companies as well. But I think they never thought of using mathematical formulas to justify. I am sure there were, but it was not well known. So, when different companies, starting with newspapers started using language models, not like this, they used to do some kind of analysis of fixed clustering or sentimental analysis, these were all stuck at that time. But again, those companies came later. I mean we had a very small number of companies who were taking people from Statistics background. So, at the same time, there were tech companies where it was easy to get a job if you just knew programming. So, I managed to learn some kind of programming in those couple of years and that's the only reason I did that. When I was doing graduation at Xavier's, I had a different idea. I wanted to pursue towards PhD and teaching. But by the end of the third year, I realised I had to go a little bit fast, that's the reason I decided to take up the applied part of Statistics. Over time, I realized that there were companies like GE CAPS, GLOBAL or GENPACT, that came up. PFIZER and other drug companies also started hiring people from the Statistics department. So that motivated me towards joining a company which is using Statistics as a tool. Then one day I saw an ad in the Times of India, where there used to be a supplement for the job market, I saw SaaS, and because by that time we had already used SaaS,

they were also present in India and then I realized "Okay, let's get in that company". Since I had my programming background, I thought it would be easier for me to merge to the statistical side as well. So, that is how it all went. And yeah, as I said I wanted to work in this domain, so, I had couple of failures or I would say these are like turning points for me because of that programming thing that I did in a software company that helped me in the long run. I mean probably that is one of the reasons why I am currently where I am. It's mainly about how you need to have your plan and the grit to get there.

Q7. What do you consider to be your greatest achievement in life so far?

Sir: Yeah, so, as I was telling you when I was in IIT, these big companies, Microsoft, Accenture, they used to come, but they never opened for us. So, at today's date, when I am working for Microsoft in their headquarters for an electricity division, where my models are there to secure their cloud, and which is helping the customers so that they can protect their domain and their data, that is definitely a nice thing I can cherish, for sure. And that is what I could say on the professional side. And on my educational side, getting through St. Xavier's College was one of the days which I really cherish. As I said, when I gave that letter to Dr. Amit Ghosh and he asked me to attend the class the next day, that was a turning point for me in my life, I would say. That was a turning point, and now, at this point, I am also cherishing what I am doing at Microsoft.

Q8. How much of what is taught at the undergraduate level is, in your opinion, relevant for later career pursuits?

Sir: The first thing is at least at that time the kind of course or the materials we went through at our undergrad level, I used those things even in my postgraduation. Some of it, even in IIT Bombay. Our course was like a mix of Computer Science subjects and Statistics and also Mathematics. For the Statistics part, in my opinion, we were far ahead in that curve coming from St. Xavier's or Calcutta University. So, it was never an issue with the Statistics part, at least for people who joined from West Bengal. So, what I can say is that, to this date, I use time series distributional fit and optimization techniques, how to estimate the weights or the parameters of the model. I at least know what is going under the hood. So, that helps me a lot about understanding the output of the model, probability. I am not saying that you have to know all those questions, but you should know that in a continuous distribution, the probability of a point is zero. So, those things really helped me. Together with that, since I did my Master's, during that time I learnt more about machine learning techniques and models because when we were undergraduates, other than OLS, and linear regression we were not taught logistic regression, so those things I learnt at a later stage in life during master's, in India and the US. Also, the algorithm data structures that helped to keep the footprint minimum, efficient programming models and kinds of stuff. So, I would say that Linear Algebra is not going to leave you any time soon and that Time Series, if you are interested in temporal data, then obviously is another learning you should know. Let's say you want to go towards the marketing side, where a lot of testing happens, then multivariate testing will be important. Even in modelling also, just because we have a lot of

data nowadays, we do not care about p-values anymore. But in earlier days, the p-value was important because we didn't have that much data and that is why we used to say that 50 observations were big data, it's a large number of data points, which you will laugh at today's date. So, that is where the p-value was important because to check the sample, to check the p-value whether it is making sense, whether I am estimating this parameter, whether the p-value of the parameter is significant or not, those kinds of things were important those days. But in today's date, in my opinion, and I know many can differ in this statement, nowadays people do not give much importance towards the p-value, it is all about fitting.

Q9. What is your take on pursuing Data Science as a future career?

Sir: If you see those deep learning sites, generally there are three profiles. Data science is used a little bit, it's a convoluted term. You will find there is a data analyst, a data scientist and an ML engineer/ML scientist. Broadly there are three categories. Data analysts are mainly about what they do, they try to analyse the data to understand the trend, or they generally get a job mainly in the marketing or sales domain where they do some kind of forecasting or testing right or wrong, that kind of stuff. On the other hand, data scientists generally create those models, use data tabular, unstructured different types of models, and deep learning stuff. They are more on the API side, they can also write libraries or create their models which are not available as an open source, and sometimes they have to write those things. But, data analysts, definitely they are more on the API side. They call APIs, create those outputs, interpret them; more towards the business side. They interact more with the business side and explain the data. Whereas data scientists and ML engineers are more towards the product side. Maybe it is wrong, but whatever I have understood from my career I am telling you that. Let's say you have a product, like Zomato, or Swiggy. How to allocate whom to which location based on the demand? There is software working under the hood, and that software is built by the ML engineers and software engineers but the POCs or the Proof of Concept or initial study, what kind of model could be fit or how can we improve the model and stuff, those things can be done by a statistician or a data scientist as well. Obviously, there are computer scientists here who also play a crucial role here.

Q10. Which roles as a data analyst, in your opinion, has the brightest prospects?

Sir: In my opinion, ML engineers are paid the most. Generally, they come from a computer science background, they do low-level coding. Let's say there are GPUs like ChatGPT, how to code the lower level ChatGPT in a GPU, probably me or statisticians won't be able to do it. There can be exceptions, people who are very much inclined towards those domains. I am not saying that it is not possible, it's just less likely. Those are the ML engineers who can help you to create the model framework, the pipeline, and the heavy work on the modelling side. Hence, they are paid the most. Data scientists and data analysts are very close by. Data analysts, based on which company you are joining, lots of time they say data analysts, but they are actually doing data science work. Based on demand, both of them are equally important. If you do a test, I think, at least here in the USA, I could see that data science is parallelly a bit more. I mean people who are interested in working in the core rather than working with the business side,

salespeople or marketing people, it's just a choice.

Q11. Which of Statistics, Mathematics, Computer Science, in your opinion, is a preferred background for navigating to Data Science?

Sir: So, in the Math department, they are also aware of all the math you are doing, mainly the linear algebra part, graph theory part which helps in the network analysis, subgraph anomalies - those kinds of things, Hub and spoke model, the network or influencer model. Those are things a mathematician as well as a computer science student can do. Computer science people are good at algorithms, so they know how to store data, how to use minimum storage, and minimum time to execute one model. They will be good at that. They are writing the APIs. On the other hand, Data Science people know exactly which API to call. Whether a T-test or a Z-test we should do, that you will know and probably a computer science person is going to overlook. You know the underlying distribution, like if it's a square of normal or chi-square, why we should have a chi-square model and not a normal distribution, those are things a data science person or a statistician can do. I would say if you were interested more towards the programming side, yeah you can go and bring an ML engineer for that, it's all about mainly coding, but coding to create the models. Data science side, statisticians will be much ahead of the curve when it comes to data analysis where you know what kind of test/model you are going to build. How the importance of a feature in a model you are going to measure; will help you. Mathematics or Physics people, definitely have similar kinds of things; they also do regression models and stuff. But think about it, they are doing it as a tool, they are not learning in-depth the way you have been doing for 3 years to understand all these concepts. Linear algebra is not going to leave you. Rather I would say that embrace it, and that is how you can show your value. I started talking about linear algebra, probability, these different kinds of models, testing, hypothesis testing-they are going to be your bread and butter. And programming, if you are good in C++ that will be awesome, you can then definitely move towards machine learning engineering. But nowadays, many people are writing more in Python. But believe me, Python is not that fast. People are working in Scala. There are other languages, Go and all that stuff. So, yes, up to you, whether you want to keep programming as your weapon or more about statistical knowledge, different kinds of algorithms to fit models, if that is your weapon.

Q12. What additional skills do you recommend to a student of Statistics for a career in Data Science?

Sir: Be aware of what is going on lately. The large language models are so huge, that they went through all the data on the web till 2021 and somehow, they managed to decipher the grammar of a language. So, that is where it is coming from. It is not memorizing all the data. The model learns the structure and the grammar, and it is there. Whenever these things come, go, and check. I usually prefer medium but take it with a pinch of salt all the medium posts are not of high quality. You must figure out whether it's right or wrong and you will have that knowledge. Coming from Statistics, you will be able to figure out whether it's right or wrong, whatever the person is writing. Even on YouTube, there are a lot of courses online. Even if you see Chennai Mathematical Institute (CMI), I think there is a person from Xavier's who is a professor over there. There are lectures by him which are free. I didn't learn modelling in my undergrad or even in my postgrad, but I

asked him to share slides on that domain and I thereafter tried to learn the basics of it. I reached out to my classmate to learn that. His name was Sounak Chakraborty. He was kind enough to send it to me and said, "Read it, you will understand." Reach out to us, we are always there to help you out. So, what I mean is that be aware of the latest things that are going on in the field and you have to be a master in it, at least you should know what is going on and how that thing is happening. In an interview, they won't ask you to write the equation or the structure of ChatGPT, but they might ask what do you mean by 'attention'. I am just giving an example. In deep learning how the backpropagation works, how those weights are initialized, how they are updated, how attention comes in the picture to update? I also do courses from Coursera. In Indian money, it's probably a lot. It is a good source; I generally prefer it over LX and Udemy. Coursera has more options, and you can learn. I don't know how much time you would get after doing your coursework. If you still have time, do it during the summer vacation, and squeeze it in. For example, vision is a very important thing; with prior knowledge of something you are trying to predict something in the future. There is other stuff, something called online learning - you keep on learning and your model is evolving and that way the model is embracing the latest behaviour. Then, there is active learning which is basically that you start with a small amount of labelled data, and use it to do the inference and send those data points which the model was uncertain about and send it to the labellers, labellers are going to use that back and remodel. These are techniques and for these, you don't have to know big theories and stuff. Believe me, it's a bit hyped as well. You just keep your mind open underneath it is a model going on with different pipelines.

Q13. Name some companies you would recommend for job opportunities?

Sir: It depends on how India's economy is doing because if you see the big companies are not as good. In India, you can get work on statistics. Mainly I would say begin with start-ups. I'm pretty sure the job openings there would be with the name of data science and statistics people can also apply there. There shouldn't be any discrimination because of that. If statistics people are more inclined towards pursuing PhD that is a different thing, but some would be interested in joining a job. You are not behind in any way from data science work. You have some theoretical knowledge but if you know the programming thing it shouldn't be a problem. I would suggest you for start-ups, there are a few companies like EY and Accenture, these consulting companies. They also have these data science practices where they will go to the client and solve their problems. So, keep your eyes open for those companies as well, they are good to start your career with. You will get more exposure from the several types of problems you will face with the clients.

Q14. Looking back, how much impact has St. Xavier's College had on your career?

Sir: After undergrad, I moved more towards the applied side. I still believe my base, or the fundamentals, the background, the groundwork have been done at Xavier's. So, all the teachers and professors there, taught us all the different subjects, like linear algebra and estimation theory were taught by Indranil Mukherjee. I really love that person. I call him Indranil da and he was awesome in teaching, at least on the linear algebra side which I can still remember and I'm using it. And then, Professor Amit Ghosh, who taught us initially the probability

and it's very important these days. Then in today's date, the market index and stuff which at that time was called Official Statistics, I guess. So, you understand that helped me in my Economics class to understand the different indices, and now if you're interested in stocks, that will definitely help. Then, Dr. Goon, though I was never a very good student in his class, but yeah, Dr. Goon helped me to understand probability, which I have to use in my daily life. So, I am really grateful to him. Then, there was another professor called DG, who used to take our distribution, fitting, and all those things and we learnt a lot. And believe me, I'm using those basic things as well because in professional work you won't find Nobel Prize problems happening there, rather it's more about very simple problems and people will be happy with simple solutions. And then obviously, Surupa di and Surabhi di took the designs of experiments, which I don't use today, but when I was part of a marketing company, I did that. So, St. Xavier's College gave me a firm impression on my life. And then, all the students, my batch mates, coming from different places, different schools and the intellectual levels that they came up with, helped me to make some good friends. And yeah, I still believe my base was done in the undergrad which I'm still reaping the benefit of today's date.

Q15. How in your opinion should a student cope with the ups and downs in life besides career and academics?

Sir: Yeah, I think I mentioned earlier as well, that these ups and downs will happen. These are small glitches in the bigger picture. So, keep your plan, whatever you have planned for, work towards it. Even if it's a small fraction of that today, do it. You'll feel satisfied that – "Okay, I did something towards that goal. Maybe it's a delta, but at least I'm going towards that goal." So, don't take these smaller failures so seriously; rather, I would suggest you fail now rather than at a later point in life. So do experiments, do take risks, it is completely fine. You are quite young. You can take risks. Do it, it's fine. Fail, it's fine. These things I have also learnt when I moved to this country. In our society, I think failure is looked down upon. It is not here. They encourage people to fail. I mean, when I was in Xavier's this winter, I was meeting the people of the data science group, I guess. I was telling them that my kids, go to the school and they're encouraged to take or do whatever they want. And the failures will come, they know that. But they are not discouraged by that. You would see that people always talk about, "Hey, in this country, everyone is A+." There's a reason behind that because it is just, they want to encourage them. It is not about the A that matters, it's the encouragement. So, maybe you are not doing that great. But if you get encouragement, who knows, you will pick that up later in life and you may prosper. So, help your friends also. Encourage them to take it positively and move forward. In the long run, it doesn't matter, those small failures. When you see after 20, 30 years, you will realise that probably that failure was needed and that made you a changed person. So, don't be afraid of failures.

FROM FIGURES TO FORTUNE: COMPARATIVE ANALYSIS OF BANKING SECTORS USING GAMMA REGRESSION MODEL

Amrita Nath[1], Poushali Dutta[1], Sreyasi Dey[1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

Introduction

The financial sector plays a critical role in the global economy, serving as the backbone for economic growth and development. Banks are a vital part of this sector as they manage vast amounts of capital and provide essential services such as giving loans, investments, and risk management. This project aims to analyse financial data from Indian banks to uncover patterns, trends, and relationships that impact their profitability, stability, and growth. Through statistical analysis, we want to have a comprehensive understanding of the banking industry's current scenario and offer valuable insights for the improvement and betterment of the industry.

Data Description

In this project, we will carry out the analysis on the bank performance data for the financial year 2023-2024. The data consists of information on 45 banks of India across three different sectors: Public sector, Private sector, and Small Finance bank. We have extracted the information on the variables 'Deposits', 'Total assets', 'Net NPA', 'Total income', 'Total expenditure', 'Net profit', 'Credit deposit ratio', 'Investment deposit ratio', 'Return on assets', 'Capital adequacy ratio' and 'Net NPA as percentage to net advances' for the year 2024. Although 'Operating expenses' is a part of Total expenditure, we have included it as well, to calculate efficiency ratio. All these variables are continuous in nature. So, we can measure them using ratio scale.

SOURCE: https://www.kaggle.com/datasets/lordpatil/indian-banks-key-business-statistics (October 2024 version)

Graphical Representations

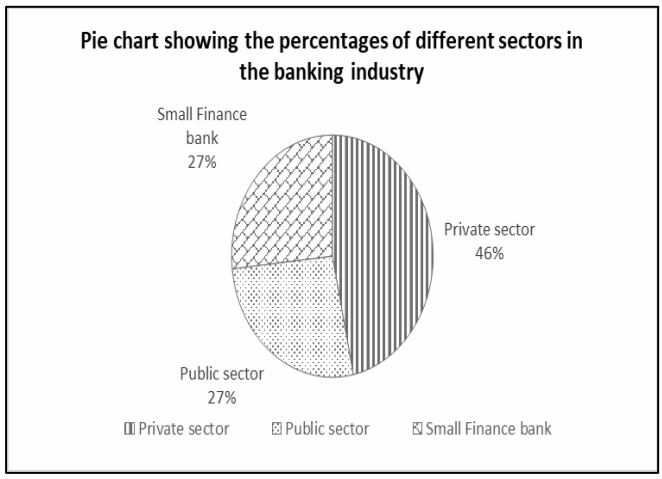


Fig 1.1

The pie chart shows that the percentage of private sector banks in the banking industry is higher than the other two sectors.

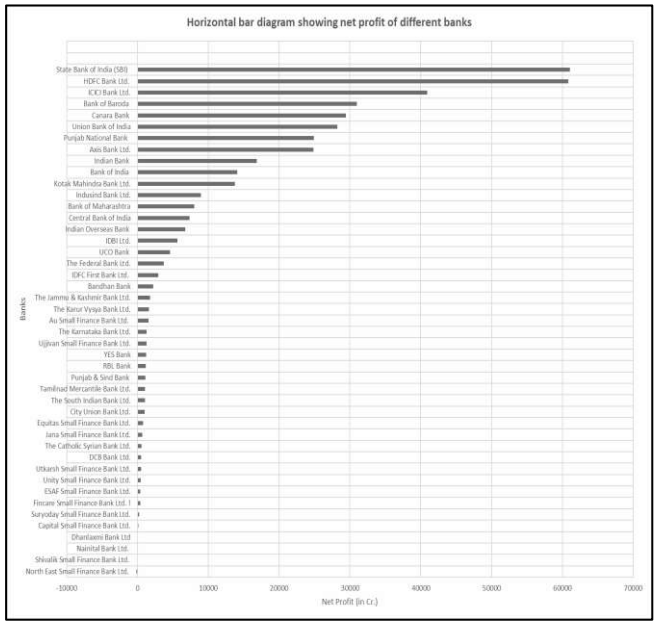


Fig 1.2

Here we have plotted a horizontal bar diagram showing net profit of 45 banks on the basis of data of 2023-24. From the plot we can say that according to net profit, the top 5 banks are State Bank of India (SBI), HDFC Bank, ICICI Bank, Bank of Baroda and Canara Bank.

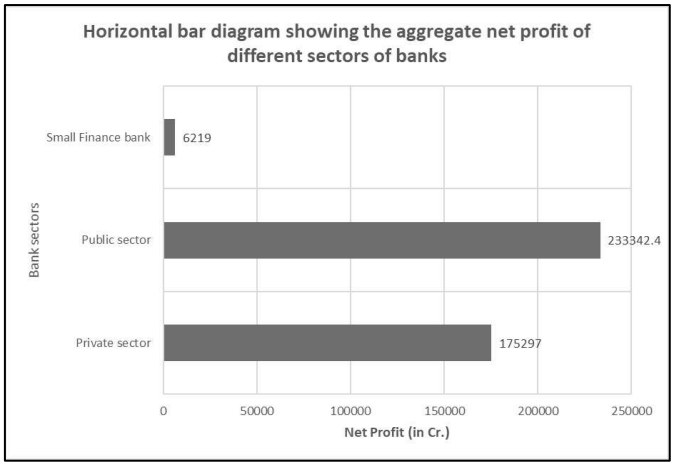


Fig 1.3

We have drawn a horizontal bar diagram of net profits across different sectors of banks. From the above graph, we can observe that the public sector banks make the most profit. Also, the small finance banks are making much lesser profit as compared to the public and private sector banks.

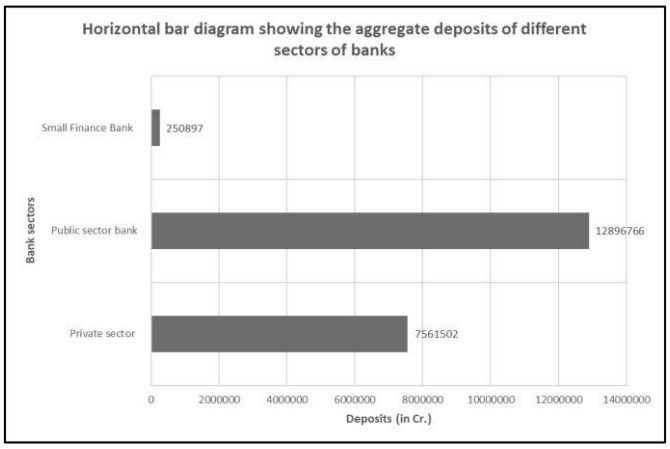


Fig 1.4

We have drawn a horizontal bar chart which shows that public sector banks have the most deposits. Also, the amount of deposits in public sector banks are very high as compared to private sector and small finance banks. This indicates that customers have more faith in public sector banks.

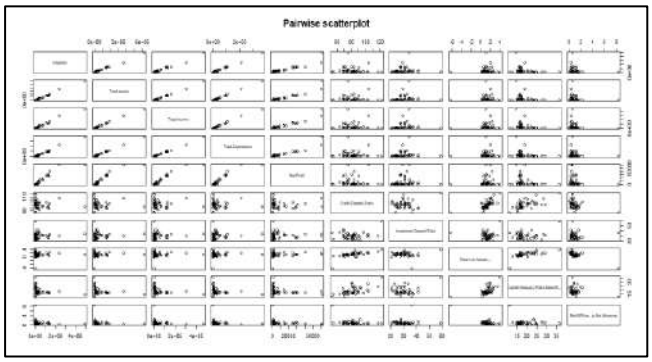


Fig 1.5

Next, we have created the scatterplot of pairs of different variables. From the scatterplot, we can interpret the correlation between the different pairs. We are interested in the factors which affect the Net profit of the banks. Here, we observe that Net profit is positively correlated (high) with Deposits, Total assets, Total income, and Total expenditure.

<u>Boxplot of Net Profit across different sectors of banks</u>

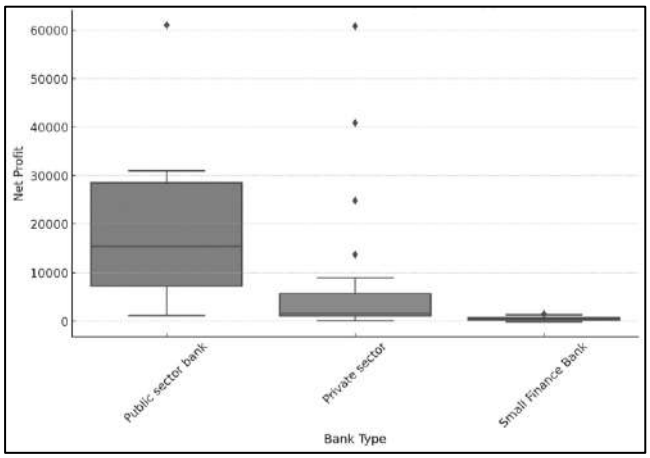


Fig 1.6

From the boxplot, we observe that Net Profit varies significantly over the different sectors of banks. Median Net Profit is higher in Public Sector banks compared to other sectors. Outliers are present in private and public sector banks indicating some banks have exceptionally high Net Profit.

We are interested in studying the relationship between Net Profit and different sectors of banks. We denote Public Sector Bank by 1, Private Sector Bank by 2, and Small Finance Bank by 3.

GAMMA REGRESSION

In classical linear regression models, we assume that the response is normally distributed. To observe the distribution of Net profit, we plot its histogram.

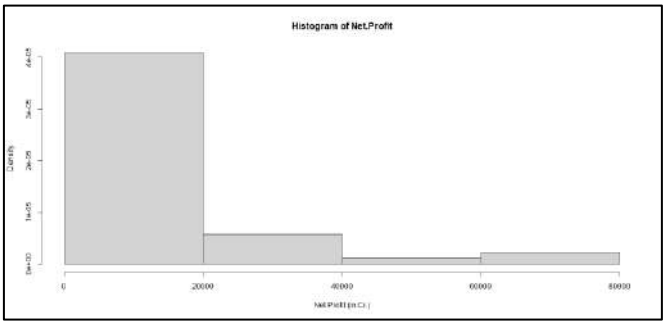


Fig 1.7

From the histogram, we observe that the distribution of Net Profit is positively skewed. Hence, we use generalized linear models, assuming that Net Profit follows a Gamma distribution.

Let Y denote the response variable "Net Profit" and x_1, x_2 denote the predictors "Total assets" and "different sectors of banks" respectively.

Here, x_2 is a categorical predictor consisting of three categories: Public Sector, Private Sector, and Small Finance Bank. Thus, we cannot incorporate sectors of banks using a single regression coefficient in the model. This is because the effect on the model when one moves from category 1 to category 2 of sectors might not be the same as when one moves from category 2 to category 3 of sectors. In such a case, we define two new predictors x_{22} and x_{23} as follows,

$$x_{22} = \begin{cases} 1 \ , & \textit{if the bank is a Private Sector Bank} \\ 0 \ , & \textit{Otherwise} \end{cases}$$

$$x_{23} = \begin{cases} 1 \ , & \textit{if the bank is a Small Finance Bank} \\ 0 \ , & \textit{Otherwise} \end{cases}$$

Note that, here we have considered two predictors: Total assets and Sectors of the bank. To check if multicollinearity exists, we calculate the Generalized Variance Inflation Factor (GVIF). Generally, GVIF values less than 5 implies low multicollinearity among the predictors.

```
                GVIF df GVIF^(1/(2*df))
Total.assets  1.2158  1          1.1026
Bank.Type2    1.5931  1          1.2622
Bank.Type3    1.7352  1          1.3173
```

Since here the values are less than 5, we can say that the effect of multicollinearity is not so pronounced.

We assume Y ~ Gamma ($\boldsymbol{\alpha}$,p), where p>0 is the shape parameter and $\boldsymbol{\alpha}$>0 is the scale parameter. The density of Y is given by,

$$f(y)= \frac{\alpha^p}{\Gamma(p)} e^{-\alpha y} y^{p-1} \ , \ \alpha>0, \ p>0$$

We have observations on n=44 individual banks (We omit the observation corresponding to North East Small Finance Bank Ltd. because its Net Profit is < 0).

We generally assume the shape parameter to be constant, which controls the skewness of the distribution. Therefore,

Y_i ~ Gamma ($\boldsymbol{\alpha}_i$,p) for all i=1(1)44

Let $E(Y_i)=\mu_i$ and $\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_{22} x_{22i} + \beta_{23} x_{23i}$, which is a linear function of predictors.

In case of Gamma regression, the canonical link i.e. the inverse link does not preserve the range of η. So, the log link is generally used which is given by g(μ)=ln (μ) where ln(.) is the natural logarithm.

The Gamma regression model is given by,

$$\boldsymbol{\ln(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_{22} x_{22i} + \beta_{23} x_{23i}, \ \forall \ i=1(1)44}$$

Using the GLM function in R, we obtain the following summary table.

```
Call:
glm(formula = Net.Profit ~ Total.assets + Bank.Type2 + Bank.Type3,
    family = Gamma(link = "log"), data = data)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.327e+00  2.791e-01  29.835  < 2e-16 ***
Total.assets 1.170e-06  1.208e-07   9.682 4.86e-12 ***
Bank.Type21 -7.354e-01  3.054e-01  -2.408   0.0208 *
Bank.Type31 -2.023e+00  3.677e-01  -5.501 2.38e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.6428164)

    Null deviance: 131.657  on 43  degrees of freedom
Residual deviance:  45.312  on 40  degrees of freedom
AIC: 817.2

Number of Fisher Scoring iterations: 20
```

Fig 1.8

Interpretations:

- The intercept value (8.327) is expressed on a logarithmic scale. In the model, we are using a log link function. To interpret it in terms of net profit, we need to exponentiate the value: $e^{8.327} \approx 4124.74$
  This implies that the predicted average net profit for public sector banks, when total assets are zero, is approximately **₹4124.74 cr**. Economically, total assets are unlikely to be zero for a bank. This interpretation is absurd, possibly due to some errors or presence of outliers.

- Positive coefficient of total assets indicates that larger total assets are associated with higher net profits. For a unit increase in Total Assets, the logarithm of expected Net Profit increases by 1.170e-06, when sector is fixed at category 1, i.e., public sector bank. The coefficient is significant since p-value is very small (at 5% level of significance), which indicates that Total Asset is a significant predictor of Net Profit.

- For private sector banks (**β**22), the log of average Net Profit decreases approximately by 0.7354 as compared to public sector banks, when total assets are fixed at a particular level. This indicates that private sector banks have lower net profits compared to public sector banks, assuming the other variables are constant.

- For small finance banks (**β**23), the log of average Net Profit decreases approximately by 2.023 as compared to public sector banks, when total assets are fixed at a particular level. This indicates that small finance banks have significantly lower net profits compared to public sector banks, assuming the other variables are constant.

Goodness of Fit

After fitting the model, we obtain the fitted values of the data using fitted( ) function in R. Now, we plot the observed and the fitted values on the same graph.

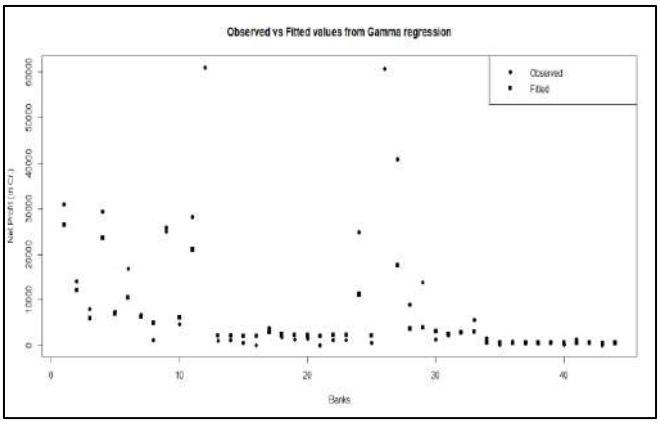


Fig 1.9

From the above plot, we note that the observed and fitted values are more or less close to each other except a few extreme values. Moreover, from the fit we observe that, Dispersion parameter: 0.6428, which is the estimate of the dispersion (or variance) in the Gamma distribution, indicating the spread of the Net Profit data around the fitted model.

Null deviance: 131.657 (with df 43)

Residual deviance: 45.312 (on df 40)

As we include the variables “Total assets” and “Sectors” in the model, the deviance decreases indicating that the variables have significant role in predicting the response variable. The reduction in deviance (from 131.657 to 45.312) indicates that the model explains a substantial portion of the variation in the dependent variable, indicating that our fit is moderately good.

EFFICIENCY ANALYSIS

The efficiency ratio shows how well a bank manages its cost. A low efficiency ratio is better as it indicates that the bank is spending less to generate income.

Efficiency=Operating Expense/Total Income

The top 5 most efficient banks are as follows:

Bank	Total income	Operating expense	Efficiency
Canara Bank	127654.4	26119.79	0.204613
Bank of Maharashtra	23492.56	4814.38	0.204932
HDFC Bank Ltd.	307581.6	63386.01	0.206079
Union Bank of India	115858.2	24439.96	0.210947
Bank of Baroda	127101.3	28251.68	0.222277

From the above table, we can observe that banks which have high net profit like SBI, HDFC Bank, ICICI Bank, Bank of Baroda, Canara Bank may not always be efficient in managing their expenses. We observe that among the top five efficient banks, four of them belong to the public sector. Among the private sector banks, HDFC has the highest net profit and is the most efficient.

Conclusion

In this project, we compared the performance of different sectors in the banking industry and observed that public sector banks dominate the industry with most deposits, highest efficiency, and highest profit. We also analysed the effect of different sectors and total assets on the profitability of banks. The Gamma regression model provides a clear understanding of these relationships, offering valuable insights for improving decision-making and strategy in the financial sector. An efficiency indicator like efficiency ratio is also important for sustainable growth. The findings highlight how statistical data analysis can guide financial institutions to make better decisions, navigate challenges, and enhance their performance.

References

1. Indian Bank Association: https://www.iba.org.in/depart-res-stcs/key-bus-stcs.html
2. Pervez, A., Ali, I. Robust Regression Analysis in Analysing Financial Performance of Public Sector Banks: A Case Study of India. *Ann. Data. Sci.* **11**, 677–691 (2024). https://doi.org/10.1007/s40745-022-00427-3
3. Ford, C. 2022 "Getting Started with Gamma Regression" UVA Library StatLab https://library.virginia.edu/data/articles/getting-started-with-gamma-regression

THE IMPACT OF AI: DELVING INTO LARGER LANGUAGE MODELS

Anindita Bag[1]

M.Sc Data Science, St. Xavier's College (Autonomous), Kolkata

A Large Language Model (LLMs) is a kind of artificial intelligence (AI) system that is able to produce text approximately resembling human dialogue or writing and is also dependent on various amounts of data. These systems are trained on massive data sets using advanced machine learning algorithms to learn the patterns and structures of human language. Large language models are becoming more and more significant in various fields such as natural language processing, speech recognition, text and code generation, and machine translation.

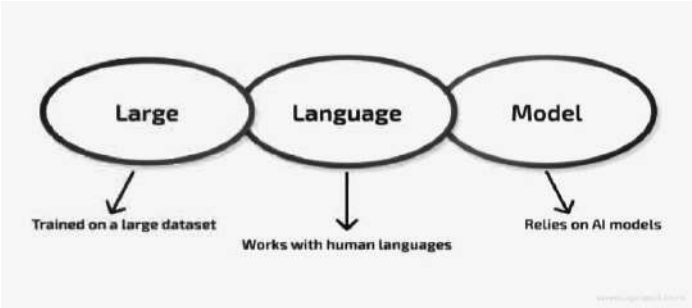


Fig 2.1

LLMs have found their way into many aspects of our daily lives, often in ways we might not even realize like when we perform tasks like:

- Translate English to Bengali
- What is the expected weather in Shimla in December
- Summarize this essay in 100 words
- Generate a tagline for a fruit juice brand that is focused on sustainable and organic farming

All these are applications of LLM.

BRIEF HISTORICAL BACKGROUND AND DEVELOPMENT OF LLMS:

- **1950s–1990s**

In the instance of translating a sentence from better English into any other language that we desired, they began regulating hard methods around the languages to be used. However, this approach comes with some barriers and works only on some systems that have been trained about.

- **1990s**

Language models begin evolving into statistical models and language patterns start being analysed, but larger-scale projects are limited by computing power.

- **2000s**

The internet changed the landscape of the functioning language models making the amount of training data grow massively.

- **2012**

GPT(Generative Pre-trained Transformer) was an architecture that was brought to life through deep learning sets and large dataset implementations.

- **2018**

The introduction of BERT (Bidirectional Encoder Representations from Transformers) by google was another major advancement which changed the architectural components of the LLM drastically.

- **2020**

A new language model specializing in language utilities enhanced based activities and communication named GPT-3 was developed with 175B parameters and set the new benchmark.

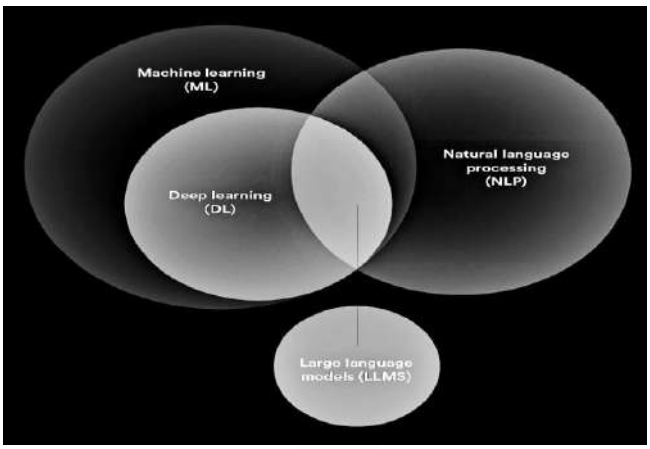


Fig 2.2

- **2020**

GPT-3 and other comparable models undergo a significant transformation with the introduction of ChatGPT, it shows incredible growth in the model's engagement rate.

BUT WHY ARE LLMS NOW MAKING HEADLINES?

The most outstanding aspects of AI language modelling have progressed thanks to some recent developments within generative AI:

• **Advancements In Techniques:** Training of these models has come a long way because of the development of new and improved techniques that have been implemented over the previous years; this has greatly enhanced earlier performance levels.

• **Increased Accessibility**: With the introduction of ChatGPT, almost anyone with an internet connection has been able to easily use one of the best language models available today and with great interfaces. This left everyone in awe of the incredible progress made with LLMs, as they were previously only made accessible to resourceful researchers or experts with extensive technical knowledge.

• **Growing Computational Power**: GPUs and even better data processing methodologies have boosted the performance of LLMs by enabling researchers to construct much larger models which boost performance.

• **Improved Training Data**: The ability to acquire and assess large sets of data has progressed as well, and this progress has translated into improved model performance.

At their essence, LLMs are deep neural networks trained on vast language data using self-supervised objectives. Their transformer -based architectures allow parallel training at gigantic scales. Today's leading models boast parameters numbering in hundreds of billions, with training datasets reaching terabytes in size. Through unsupervised pretraining, LLMs learn rich linguistic representations and contextual relationships between words. This imbues them with broad language skills like reading comprehension, conversational ability, translation and more - without any task-specific training.

To apply these general skills, LLMs can be **fine-tuned on domain targeted datasets**. This minor additional training enables them to excel at applications like question-answering, summarization or code generation. Alternatively, models may retain their wide-ranging abilities while optionally being prompted or constrained for safe, beneficial use. With their unprecedented scale and pretraining, LLMs have transformed the landscape of natural language capabilities.

TYPES OF LARGE LANGUAGE MODELS

There are several significant types of LLMs, which utilize different architectures and training techniques:

- **Autoregressive Models:** These were some of the earliest LLMs, including OpenAI's GPT models. They are trained to predict the next word or token in a sequence using previous context. Generating text is done sequentially, one token at a time.
- **Autoencoding Models**: Examples are BERT, T5, and BART. These models encode the entire input sequence into a latent representation and then decode it back into the original sequence. This allows them to be fine-tuned for various downstream NLP tasks.
- **Encoder-Decoder Models**: Like GPT-3, these models have separate encoder and decoder components. The encoder ingests the input text while the decoder generates the output text. This architecture provides more flexibility.
- **Sparse Models**: To reduce computational costs, some models use sparse representations and attention mechanisms. Sparse Transformer is one example of this approach.

As LLMs grow in size and complexity, new architectures continue to emerge. But most state-of-the-art models use the Transformer architecture in some form.

HOW DO LLMS REALLY WORK?

At their technological core, LLMs rely on the transformer architecture first introduced in 2017. Transformers represent the current state-of-the-art for language tasks by tackling long-standing challenges with prior approaches like RNNs.

- **Stacked Self-Attention Layers**: Transformers consist of stacked encoding-decoding layers containing two sub-layers: multi-head self-attention followed by point-wise feed-forward networks. Self-attention mechanisms relate different positions in a sequence to compute representations for downstream processing. This attention-based approach allows relationships between all parts of an input sequence to be mapped simultaneously in parallel. It also avoids issues with long-term dependencies that recurrent models struggle with.
- **Pretraining Through Self-Supervision:** LLMs are initially trained on massive corpora using self-supervised objectives that require no human labelling. Chief among these is the masked language modelling (MLM) task - predicting randomly masked tokens based on surrounding context. This pretraining endows models with broad language intuitions applicable across domains. Once learned, these representations serve as highly useful starting points for downstream optimization through task-specific fine-tuning or continual self-supervision.
- **Gradual Progress Through Scale**: Scaling up all aspects of language models - from their architecture depth and breadth to their training datasets and compute resources - has consistently improved capabilities. Recent paradigms like InContext Learning allow massive models to be queried through prompts for broad applications.

With their powerful yet aligned properties, LLMs have become indispensable tools paving the way towards language-focused artificial general intelligence. Let's delve deeper into some of their inner workings and impactful uses.

KEY COMPONENTS:

Under the hood, LLMs leverage key architectures and components:

- **Transformers:** Most modern LLMs use the Transformer, a neural network architecture based on self-attention mechanisms, which captures long-range dependencies in text.
- **Embedding Layers**: These convert vocabulary tokens into dense vector representations that encode their meaning.
- **Context Windows**: The fixed-length text snippets are input to the model during training and inference. More extended contexts allow modelling longer dependencies.
- **Parameters:** The trainable weights that store the model's learned knowledge, which for large models can exceed hundreds of billions of parameters.

WIDESPREAD APPLICATIONS OF LLMs:

- **Chatbots and Virtual Assistant:** The primary use case of this kind of technology, organizations can use LLMs to build automated solutions for providing assistance in customer support, troubleshooting, or even simply engaging in conversations based on the user's input.

  Code Generation: Unlike existing systems which are always used to respond to queries, built in LLMs can be trained using a vast hoard of codes and examples, letting them obtain useful code snippets useful for completing natural language requests.
- **Sentiment Analysis and Opinion Mining:** Especially when dealing with voices, sentiment analysis can be a hard task to quantify. However, LLMs in combination with audio data can be drawn to emotion and opinion data, collecting the necessary data and feedback for organizations to enhance customer satisfaction.

- **Text Classification and Clustering:** The ability to categorize and sort large volumes of data enables the identification of common themes and trends, supporting informed decision-making and more targeted strategies.
- **Language Translation Only:** Globalize all your content without hours of painstaking work just by passing your web pages through high grade LLMs to gain translation around your content.
- **Summarization And Paraphrasing:** All customer phone calls or meetings can be condensed into short succinct summaries that can be read easily. If you have a long article or speech you want to summarize, LLMs can help.
- **Creation of complex prompts and LLM generation of detailed outlines:** With a clear articulation of the subject, you want to delve into, give an extensive prompt to the LLM and then have it curate a suitable outline.

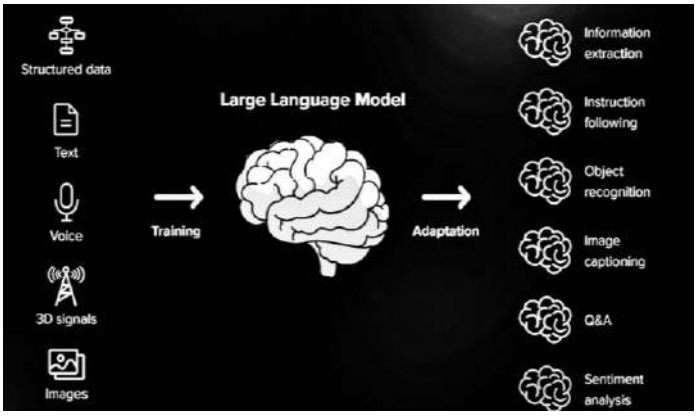


Fig 2.3

CONCLUSION:

Large Language Models have transformed AI-powered communication, providing unparalleled functionality in text creation, translation, summarization, and beyond. Their creation, fuelled by the advancement of deep learning and computing capabilities, continues to transform industries and daily interactions. Yet, as the models advance, ethical concerns like bias, disinformation, and responsible use need to be addressed. Finding the right balance between innovation and ethical use will be the most important factor in making LLMs beneficial to society with minimal risks. Thus, Large language models are transforming AI with vast potential, but addressing ethical challenges and biases are crucial for their responsible development and societal impact.

REFERENCES:

1. https://www.cloudflare.com/learning/ai/what-is-large-language-model/

2. https://www.geeksforgeeks.org/large-language-model-llm/

3. https://www.ibm.com/think/topics/large-language-models

A PROBLEM IN CAR PARKING: A RECURSIVE APPROACH

Anubhav Hazra [1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

Problem

Let us consider a scenario. One of your relatives is throwing a party to which you're invited, at their home, which is situated quite far away from your home, as a result of which, you are to go there driving your car. Let us say, that the route to your relatives' home is a one-way route and you need to park your car in one of the parking slots, arranged along the road on which their home is placed.

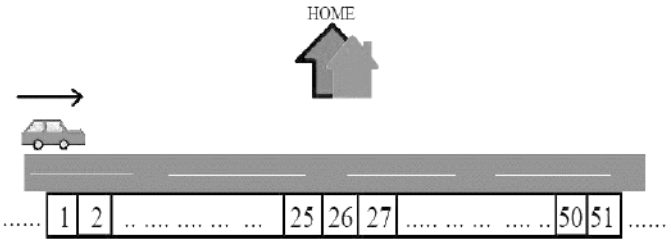


Fig 3.1

As shown in Fig 1, there are a large number of parking spots on either side, spot "26" being exactly opposite to, and thus the closest to their home. For our understanding, we've only numbered the closest 25 spots on either side of their home in the figure. As you approach your relative's home from the left, suppose you find the 1st spot vacant, but the problem is that it's too far away from the home. You would ideally prefer a closer spot so that you would have to walk less. The catch in this scenario lies in the fact that when you're with your car in front of a certain spot, you don't get to see whether the succeeding spot(s) are occupied or vacant. The only way you would figure that out is when you'll drive to that spot. Here, arises the difficulty in deciding whether or not to park in a vacant spot that we find early, or to go further to find a "possible" vacant spot that is closer to the home, solely to have a shorter distance to walk between the home and the parking spot. The greed to find a closer spot may backfire as we can't drive backwards, considering one might want to do so as it is less likely to find vacant spots near the home. We obviously need to consider that some of the guests have already arrived and most of them are likely to have parked near the home. To summarize, our objective is to form a strategy for this problem, how to decide whether or not to park at a particular spot. (Owing to some assumed conditions discussed later). Also, note that the distance to walk between the home and the parking spot would be measured in several spots, i.e., let's say you parked your car at spot 39, the distance to walk would be = 39 – 26 =13 spots.

Approach

There could obviously be multiple approaches, such as, "onwards 26, park in the first open spot found", or like, a no-risk approach of "park in the first ever open spot found", and so on. The broad motive is to decide that, whenever we are in front of a certain parking spot which is vacant, should we park there or should we skip that and drive to further spots (hoping to get a closer spot to the home)? We will examine the situation at different parking spots, and thus draw conclusions for each spot, i.e., decide whether to park or not. We first understand this for the spots 26 and later, followed by spots to the left of 26, i.e. 25, 24, … and so on leftwards.

Formulation

Case 1: Let us say we are at spot 26, and it is vacant

The obvious approach should be if we do happen to find spot 26 vacant, it is a no-brainer to park, as the distance would be zero, i.e., it can't get better than that.

Case 2: Let us say we are at any spot after 26, i.e., 27 or 28 or …., and we find it to be vacant.

Even in this case, we must park. Notice that, skipping any of these vacant spots would only lead us to go further away from spot 26, i.e., it would only get worse if we skip. Thus, we should park if we find the spot to be vacant.

Case 3: We are at an arbitrary ith spot (i<26), which is vacant. Should we park or skip and drive ahead? How to decide?

Now, if skip the ith spot, we would be moving on to the (i+1)th spot, which is also a factor here. Whether or not should we skip the ith spot has to be decided by a simple comparison:
Let A_i denote the distance from the home if we are choosing to park at the ith spot.

And let B_i denote the expected distance of the further (i+1, i+2, i+3…) spots from the home, if we are deciding to skip the ith spot and move ahead.

Thus, if $A_i \leq B_i$, we should prefer parking at the ith spot

Else, we should drive ahead.

(i) We are at spot "25" and it is vacant.

So, A_{25} = Distance from the home if we park at spot 25= 26 - 25 =1

B_{25} = Expected distance from the home (calculated onwards the 26th spot)

Let X be a random variable denoting the number of taken spots (counted onward the 26th spot) before the first vacant spot appears. Thus, X = 0,1,2,3…. One can say that X is nothing but the distance between the 26th spot and a random spot onward spot 26. We assume X to follow a geometric distribution i.e., X ~ Geo (p), where p is the probability that a randomly selected spot would be vacant.

The probability mass function of X is given by f (x) = P(X = x) = $\begin{cases} (1-p)^x p\,, & x \geq 0 \\ 0 \quad\quad , & x < 0 \end{cases}$

Note: We assume that the probability that a randomly selected spot would be vacant i.e., p = 0.1

Thus B_{25} = E(X) = (1-p)/p = 0.9/0.1 = 9

Hence, $A_{25} < B_{25}$, thus, on average, it is a **better decision to park if we are at spot 25 if it is vacant.**

(ii) We are at spot "24" and it is vacant.

So, A_{24} =Distance from the home if we park at spot 24=26 - 24 =2

B_{24} =Expected distance from the home (calculated onwards the 25th spot)

Thus,

B_{24} = P(spot 25th is vacant)*(Distance between 25th and 26th, because we would be parking at 25th if it is vacant, as decided from the previous case) + P(spot 25th is taken)*(Expected distance onwards the next spot i.e. 26, because we would be moving to spot 26 if 25 is taken)

= (0.1*1) + (0.9 * Expected distance from the home (calculated onwards the 26th spot))

= (0.1*1) + (0.9 *B_{25}) [As calculated above, B_{25}=9]

= 8.2

Hence, $A_{24} < B_{24}$, thus, on average, it is a **better decision to park if we are at spot 24 if it is vacant.**

(iii) We are at spot "23" and it is vacant.

So, A_{23}=Distance from the home if we park at spot 23=26 - 23 =3

B_{23} =Expected distance from the home (calculated onwards the 24th spot)

Thus,

B_{23}=P(spot 24th is vacant)*(Distance between

24th and 26th, because we would be parking at 24th if it is vacant, as decided from the previous case) + P(spot 24th is taken)*(Expected distance onwards the next spot i.e. 25, because we would be moving to spot 25 if 24 is taken)

= (0.1*2) + (0.9 *Expected distance from the home (calculated onwards the 25th spot))

= (0.1*2) + (0.9 *B_{24}) [As calculated above, B_{24}=8.2]

= 7.58

Hence, $A_{23} < B_{23}$, thus, on average, it is a **better decision to park if we are at spot 23 if it is vacant.**

This calculation is a recursive process, where we keep going on substituting the expected distances in successive steps. However, notice how the values of B_i keep decreasing. The obvious question arising is, till what point would we keep choosing to park, or like, is there a certain threshold?

Further calculations:

i-th spot	A_i	B_i
22	4	7.122
21	5	6.8098
20	6	6.62882
19	**7**	**6.5659**

Well, on further continuing the calculations, we observe the threshold to be at spot "19". When we are at spot "19" and we choose to park, the distance is 7. If we do not park, and move to spot 20, the expected distance onward spot 20 is 6.5659, which is less than 7. So, if we are at spot 19, we are expected to find a further spot whose distance is less than or equal to 7. Thus, it isn't a very big risk to turn down spot "19". Note that, this threshold would change with the change in the value of p, i.e., the threshold spot is 19, when the probability of a random spot being occupied is 0.9 (or 90%).

Suppose this probability of a spot being occupied was lower, at say 70%, then we'd be more likely to find a closer spot, thus the threshold spot would be above 19, i.e., 20 or higher. On the contrary, if this percentage was higher than 90, say 97%, we'd be less likely to find a closer spot; thus, the threshold would be below 19, i.e., 18 or lower.

Thus, to sum up how the strategy looks like as a whole, (with the assumption that the probability of any parking spot being occupied is 90%, i.e., p = 0.1)

Moving from left to right, till spot 19, it is advisable to not park and drive ahead,

And for spots after 19, park at the very first vacant spot available.

Conclusion

Therefore, we have actually ended up with an optimum strategy to tackle this problem. This however, as mentioned above, would change for different values of p, regarding which we can obviously generalize our strategy as a function of p, by accordingly making necessary substitutions. This would anyway mean that we would need to know the value of p. For unknown values of p, we would have to estimate it by previous sample data. Some other interfering factors could be the time at which these samples would be drawn, i.e., there would obviously be more occupied spots 5 minutes before the party starts as compared to say, an hour before the party. Thus, we definitely need to take care of the information we would be using to formulate our approach.

Reference:

- https://www.math.ucla

CASE STUDY: APPLYING THE MONTY HALL PROBLEM TO INVESTMENT BANKING

Arnabi Sarkar [1], Sreetama Dey [1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

An Insight into the Monty Hall Paradox: Imagine a game show scenario where a contestant is presented with three closed doors. Behind one door is a brand-new car, symbolizing a highly desirable prize, while the other two doors hide goats, which are considered less appealing prizes. The contestant's goal is to select the door with the car behind it.

When the contestant chooses a door, there is initially a 33.33% probability that their chosen door conceals the car and a 66.67% probability that one of the other two doors does so. At this point, the game takes an interesting turn: the host, who knows what lies behind each door, opens one of the remaining two doors, always revealing a goat. Now, the contestant is given the option to either stick with their original choice or switch to the other unopened door.

The key question here is: What should the contestant do? At first glance, it seems like it shouldn't matter whether the contestant switches or stays since there are only two doors left, and thus, one might assume the probability is now 50% for each door. However, this reasoning is flawed. The paradox reveals that the contestant should always switch doors because doing so increases their probability of winning the car to 2/3, while staying with the original choice only yields 1/3 chance of success.

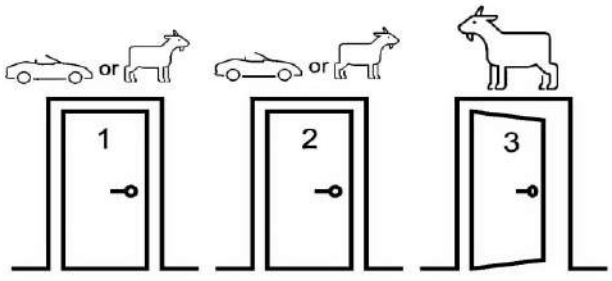


Fig 4.1: Pictorial representation of the scenario of Monty Hall Paradox

Monty Hall Paradox Table: The following table summarizes the outcomes of the Monty Hall problem for each possible initial door choice and whether the contestant switches doors or not.

Initial Choice	Behind Chosen Door	Switching Door	Outcome (Winning the Car)
Door A	Car ($\frac{1}{3}$ chance)	Door B or C (Goat)	Staying: $\frac{1}{3}$, Switching: $\frac{2}{3}$
Door B	Goat ($\frac{2}{3}$ chance)	Door A (Car)	Staying: 0, Switching: 1
Door C	Goat ($\frac{2}{3}$ chance)	Door A (Car)	Staying: 0, Switching: 1

Table 4.1: Monty Hall Paradox

Understanding Why Switching Works: To understand why switching is the optimal strategy, let's break down the probabilities. When the contestant first selects a door, they have a chance 1/3 of having chosen the car and a chance 2/3 that the car is behind one of the other two doors. The host's action of revealing a goat is crucial because it is not random—it is based on the host's knowledge of what lies behind each door.

By revealing a goat, the host effectively confirms that if the car is not behind the contestant's original choice, it must be behind the other unopened door. Thus, if the contestant's initial choice was incorrect (which happens 2/3 of the time), switching will always lead to the car. Conversely, if their initial choice was correct (which occurs only 1/3 of the time), sticking with it is the better option. Therefore, switching doors is statistically the

better strategy, doubling the chances of winning.

Investment Banking Analogy

Now, let's translate this counter-intuitive insight into the world of investment banking. Imagine an investment bank managing a diversified portfolio for a high-net-worth client, with initial investments allocated equally across three major sectors: Technology, Healthcare, and Real Estate. Each sector represents a different door in the Monty Hall problem, with only one sector likely to outperform significantly.

Initially, based on historical data and market analysis, the bank assumes that each sector has an equal chance of outperforming, i.e., 1/3 probability for each. However, as the market evolves, new information becomes available that may affect the likelihood of each sector's performance.

Market Disruption and Updated Information

Suppose midway through the investment period, new regulations are introduced that heavily impact the Real Estate sector, analogous to the Monty Hall host revealing a goat behind one of the doors. The bank now faces a strategic decision: should it stick with its original allocation or adjust its portfolio to take advantage of this new information?

In this scenario, the new regulatory environment significantly increases the risk of underperformance for Real Estate investments, shifting the odds in favour of the other two sectors. Just like in the Monty Hall problem, the optimal strategy involves reallocating investments based on the updated information rather than remaining committed to the original allocation.

Decision Analysis Inspired by the Monty Hall Problem

When the game starts, there is a 1/3 chance that any one sector (e.g., Technology, Healthcare, or Real Estate) will be the best performer. However, with the new information about regulatory challenges in Real Estate, the probability distribution shifts. The chance is 2/3 that one of the other two sectors (Technology or Healthcare) will outperform now becomes even more significant.

The bank's decision, therefore, should not be driven by initial assumptions but rather by updated probabilities based on the latest data. This is where concepts from probability theory, like the Monty Hall paradox, can be extremely useful for strategic decision-making in uncertain environments.

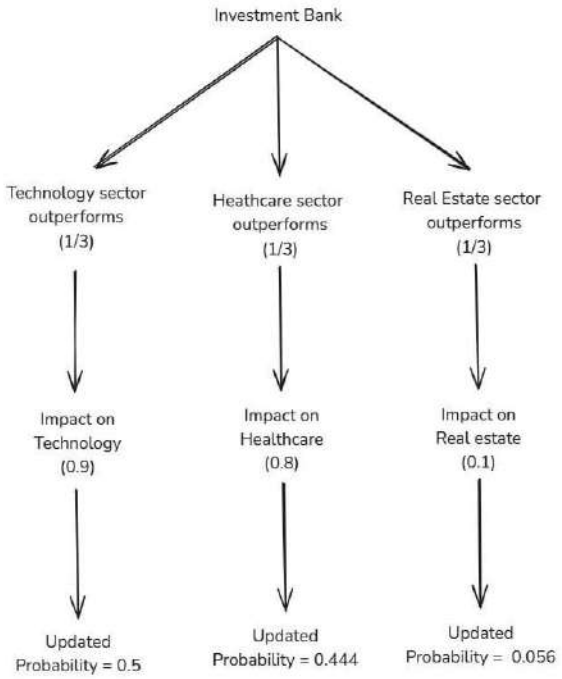


Fig 4.2: Decision Tree of the scenario of Investment Banking

Applying Bayesian Probability to Investment Decisions

To quantify this decision-making process, let's apply **Bayesian probability** to update our beliefs based on the new information. Initially, the bank assumes equal probabilities.

P(Technology outperforms) = P(Healthcare outperforms) = P(Real Estate outperforms) = $\frac{1}{3}$

Notations:

- A_1: Technology sector outperforms.
- A_2: Healthcare sector outperforms.

• A_3: Real Estate sector outperforms.

Step 2: Assigning Likelihoods Based on Impact

Based on market analysis, the impact of the new regulations on each sector is estimated as follows:

$P(B|A_1) = 0.9$ (low impact on Technology),

$P(B|A_2) = 0.8$ (low impact on Healthcare),

$P(B|A_3) = 0.1$ (high impact on Real Estate).

Step 3: Calculating the Marginal Probability

$$P(B) = \sum_{i=1}^{3} P(B|A_i) \cdot P(A_i) = (0.9 \times \frac{1}{3}) + (0.8 \times \frac{1}{3}) + (0.1 \times \frac{1}{3}) = 0.6.$$

Step 4: Calculating Posterior Probabilities

Using Bayes' theorem, we update the probabilities:

$$P(A_1|B) = \frac{0.9 \times \frac{1}{3}}{0.6} = 0.5,$$
$$P(A_2|B) = \frac{0.8 \times \frac{1}{3}}{0.6} \approx 0.444,$$
$$P(A_3|B) = \frac{0.1 \times \frac{1}{3}}{0.6} \approx 0.056.$$

Strategic Decision: The bank should reallocate funds away from Real Estate, focusing on Technology and Healthcare, which now have a combined probability of 0.944 of outperforming.

Conclusion

The Monty Hall problem teaches us that intuitive decisions are not always optimal. By continuously reassessing and updating beliefs based on new information, investment banks can enhance portfolio performance. This case study demonstrates the importance of agility in investment strategies, where using probabilistic reasoning helps maximize returns and minimize risks in dynamic market environments.

References

1. The Monty Hall Problem: The remarkable Story of Math's most Contentiuos Brain Teaser by Jason Rosenhouse.
2. Investment Banking (4th Edition) by Pratap Giri S

ESTIMATING π USING NEEDLES

Debdatta Bhattacharya[1] , Rounak Ghosh[1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

1. Introduction:

The journey of Mathematics started with the beginning of Universe. It is the root of all progress of the Human Civilization from the primitive age to the era of technology. The language of numbers has the supreme capability to define and understand every event happening in this universe. The discussion on the language of the universe becomes incomplete without the wonderful transcendental number π. It is the ratio between a circle's circumference and its diameter. The number π has wondered all Mathematicians from ancient civilization to our present time. Different great Mathematicians have discovered various fascinating methods to approximate π. One early example of this is the solution to the Needle Problem, which was proposed in 1777 by the French mathematician Georges-Louis Leclerc, Comte de Buffon. It is one of the earliest problems in geometrical probability which shows that π can be estimated using empirical data and repeatedly dropping a needle and recording its outcomes offers us with a Monte Carlo approach to approximate the value of π.

2. Buffon's Needle Problem:

The Needle problem is a wonderful question on geometric probability which uses geometric shapes, numerical integrations and real-world estimations to compute probabilities. Let's go through it and savour its taste.

Problem: Consider a vertical board with horizontal parallel lines spaced at a constant distance, 'a' apart. Suppose a needle of length 'l' (l<a) is randomly thrown onto the board. Now, we will determine the probability, 'p' that the needle intersects one of the parallel lines.

Solution: To solve this problem, we will use the concept of the polar coordinate system. Let 'y' represents the distance from the needle's centre to the nearest parallel line, and let ϕ denotes the angle formed between the needle and this parallel line.

These two quantities, y and ϕ, completely determine the position of the needle.

The problem is depicted in the Figure – 1. Clearly, the value of y ranges from 0 to a/2 (since l < a) and the value of ϕ will range from 0 to π.

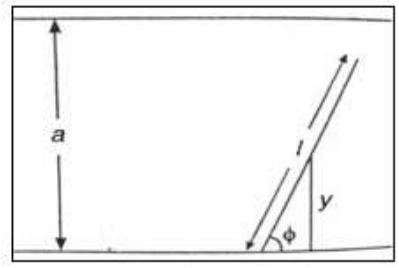


Fig 1

Since the needle is dropped randomly, all possible values of y and ϕ can be considered equally likely. As a result, the joint probability density function f (y, ϕ) for y and ϕ follows a uniform distribution. –

$f(y, \phi) = k;\ 0 \le \phi \le \pi,\ 0 \le y \le a/2$, where k is a non-negative constant.

From the Figure – 1, we can enumerate the value of y when the needle will touch the line nearer to its centre.

$$\sin\phi = \frac{\text{Height of the Triangle}}{\text{Hypotenuse of the Triangle}} = \frac{y}{\frac{l}{2}}$$

$$\Rightarrow \quad y = \frac{l}{2} \times \sin\phi$$

Therefore, the needle will cross one of the lines if the distance of its centre to the nearest line is less than $\frac{l}{2} \times \sin\phi$. So, here the event of interest can be expressed by using the inequality,

$0 < y < \frac{l}{2} \times \sin\phi$.

Therefore, the probability p for this event can be determined as:

$$p = \frac{\int_0^{\pi} \int_0^{\frac{l}{2} \times \sin\phi} f(y,\phi)\, dy\, d\phi}{\int_0^{\pi} \int_0^{a/2} f(y,\phi)\, dy\, d\phi} = \frac{2l}{a\pi}$$

If there are N number of needles of which x needles have the needle intersecting the line, then the probability can be approximated by the proportion $\frac{X}{N}$. By notation,

$$p \approx \frac{x}{N}$$

This can be written as –

$$\pi \approx \frac{2lN}{ax}$$

Now, in case of a longer needle where l>a, the calculation becomes more complex as needle can cross multiple lines. We need b (ϕ) which lies between ½ sin ϕ and a. So, by integrating the joint PDF we get,

$$p = \int_{\phi=0}^{\pi} \int_{y=0}^{b(\phi)} \frac{4}{a\pi} dy\ d\phi$$

3. Approximating π:

In 1901, Italian Mathematician Mario Lazzarini performed this experiment. He took needles of length $\frac{5}{6}$ of the width between the two straight lines and tossed 340 needles. So, the chance of the needles crossing the lines in the experiment is 5/3π (Since l/a is 5/6). Lazzarini was aiming for the value 355/113 (approximation of π), So

355/113 = 5/3 * N/x ⇒ x = 113/213 * N ≈ 1808

Hence, he counted 1808 needles which had intersected its nearest line. From these observations we can compute the estimated value of π as –

$$\pi = \frac{5}{3} \times \frac{3408}{1808} = 3.1415929204$$

This experiment conducted by Lazzarini, is a case of confirmatory bias, where if we drop 213 needles and gets 113 successful cases, in that case we can obtain an estimate of π which is accurate to six decimal places. Otherwise, we must conduct an additional 213 trails, aiming 226 successful cases and if not, then we need to repeat as necessary. Lazzarini carried out 213 x 16 attempts, making this strategy appear as his method of obtaining the required estimate. This estimate tends to the appropriate value of π as the number of needles (N) tends to infinity.

4. Other Methods:

In the short needle case (l<a), another method has been used to solve the needle problem by using geometry. For a needle of length 1 unit and the space between the horizontal lines being 2 units, we define d as the distance from the needle's centre to the nearest line and θ as the acute angle between the needle and the parallel lines. So, the needle will intersect the closest line when $d \leq \frac{1}{2} \sin\theta$. In the graph (Fig 2) given below, the shaded region is the condition of the needle crossing the horizontal line.

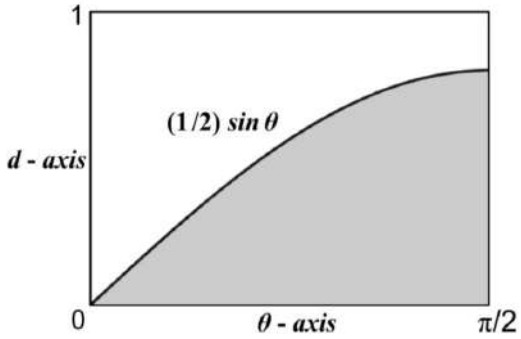


Fig 2

Area under curve = $\int_0^{\pi/2} \frac{1}{2} \sin\theta\, d\theta = \frac{1}{2}$

Probability = $\frac{Area\ under\ curve}{Area\ of\ Rectangle} = \frac{1/2}{\pi/2} = \frac{1}{\pi}$

5. Conclusion:

The solution of Buffon's Needle Problem becomes interesting due to the appearance of π , as the probability distribution function for the needle's orientation is rotationally symmetric, which helps us to find its value without any significant knowledge on Higher Mathematics. It's ability to estimate π provides a connection between randomness and determinism, all the while helping in significant developments in

statistical and computational methods. It highlights the surprising connection between probability and the value of π, a fundamental constant in mathematics.

6. References:

- Fundamentals of Mathematical Statistics by S.C. Gupta and V.K. Kapoor
- An Introduction to Probability Theory and Its Application by Feller and Wiley
- Advanced Engineering Mathematics by Kreysig
- An Introduction to Geometrical Probability by Mathai
- Lazzarini's Lucky Approximation of π by Badger and Lee
- Introduction to Mathematical Probability by Uspensky and James Victor

AN INTRODUCTION TO NEURAL NETWORKS: THE FOUNDATION OF DEEP LEARNING

Dipanjan Chakroborty [1]

M.Sc Data Science, St. Xavier's College (Autonomous), Kolkata

1. Overview

Neural networks, inspired by the human brain, are pivotal in modern artificial intelligence and machine learning. They are used for pattern recognition, classification, regression, and various other computational tasks. This article explores their structure, learning process, applications, challenges and future developments.

2. Structure of Neural Networks

A neural network consists of interconnected nodes, commonly referred to as neurons. These nodes are organized into three main layers:

- Input Layer: This receives raw data, where each neuron corresponds to a feature in the dataset.
- Hidden Layers: These layers process input data using weights and activation functions, enabling the network to learn complex relationships.
- Output Layer: The final layer provides the result, whether a classification, regression value, or probability distribution.

Connections between neurons are weighted, and these weights are adjusted through training. Activation functions such as ReLU and sigmoid introduce non-linearity, allowing the network to capture intricate patterns. The depth and complexity of hidden layers determine the model's ability to extract meaningful patterns from large datasets.

3. Learning Process

Neural networks learn through an iterative process known as backpropagation, which consists of:

1. Forward Propagation: Data moves from the input layer to the output layer, generating an initial prediction.
2. Loss Computation: The prediction error is measured using a loss function, such as Mean Squared Error (MSE) for regression or Cross-Entropy for classification.
3. Backward Propagation: The error is propagated backward to adjust weights using optimization techniques like gradient descent.
4. Iteration: This process is repeated over multiple epochs to minimize error and optimize performance.
5. Hyperparameter Tuning: Learning rate, batch size, and optimization algorithms like Adam or RMSprop influence network performance.

4. Types of Neural Networks

Several types of neural networks cater to different applications:

- Feedforward Neural Networks (FNNs): Data flows in one direction. Used for classification and regression tasks.
- Convolutional Neural Networks (CNNs): Designed for image and video processing, leveraging convolutional layers to detect spatial patterns. CNNs reduce computational complexities by utilizing pooling layers.
- Recurrent Neural Networks (RNNs): Suitable for sequential data like time series and natural language processing. Variants such as LSTMs and GRUs improve long-term dependencies by retaining past information.

- Generative Adversarial Networks (GANs): Used for generating synthetic data, such as images and text, by pitting a generator against a discriminator. These networks have revolutionized areas such as deepfake creation and image synthesis.
- Transformers: Commonly used in NLP tasks, transformers rely on attention mechanisms to process sequences efficiently. Models like BERT and GPT leverage transformers to achieve state-of-the-art performance in language understanding.

5. Applications of Neural Networks

Neural networks have driven innovation across multiple industries, including:

- Computer Vision: Object detection, medical imaging analysis, and autonomous driving. CNNs help classify medical scans and identify diseases with high accuracy.
- Natural Language Processing (NLP): Machine translation, sentiment analysis, chatbots, and summarization. Neural networks enable advanced language models to comprehend and generate human-like text.
- Healthcare: Disease prediction, drug discovery, and personalized treatment recommendations. Deep learning models assist in diagnostics and medical research, improving patient outcomes.
- Finance: Fraud detection, risk assessment, and algorithmic trading. AI-powered models analyse market trends and detect anomalies to prevent financial fraud.
- Robotics: Autonomous navigation, object recognition, and intelligent automation. Neural networks empower robots to make decisions based on sensory inputs.
- Entertainment: Recommendation systems used suggestions on platforms like Netflix and Spotify are driven by deep learning models.

6. Challenges and Limitations

Despite their success, neural networks face several obstacles:

- Data Dependency: Large datasets are needed for effective training. Limited labelled data can hinder performance in supervised learning tasks.
- Computational Cost: Training deep models requires powerful hardware like GPUs and TPUs, leading to high energy consumption.
- Overfitting: Without proper regularization, models may perform well on training data but fail on new data. Techniques like dropout and L2 regularization mitigate overfitting.
- Interpretability: Understanding how neural networks arrive at decisions remains challenging. Model explainability techniques like SHAP and LIME aim to improve transparency.
- Ethical Concerns: Issues such as bias in AI, privacy risks, and misuse of generative models raise concerns about responsible AI deployment.

7. Future Prospects

Advancements in neural networks focus on improving efficiency, reducing data dependence and enhancing interpretability. Research in hybrid models, transfer learning, and quantum computing is expected to revolutionize the field. As AI continues to evolve, neural networks will play an even more significant role in shaping technological advancements.

Key areas of future development include:

- Few-Shot Learning & Transfer Learning: Reducing dependency on large datasets by leveraging knowledge from pre-trained models.

- Neuromorphic Computing: Designing hardware architectures that mimic biological neural networks for enhanced efficiency.
- Ethical AI Research: Developing guidelines and frameworks for responsible AI development, ensuring fairness and bias mitigation.
- Quantum Neural Networks: Harnessing quantum computing for faster and more efficient neural network training.

8. Conclusion

Neural networks are at the heart of modern AI, enabling breakthroughs in numerous domains. With continuous improvements and ethical considerations, their potential remains boundless, paving the way for a future driven by intelligent systems. As deep learning continues to evolve, neural networks will shape the next era of artificial intelligence, influencing fields from healthcare to autonomous systems and beyond.

9. References

a) Neural Networks and Deep Learning by Michael Nielsen – http://neuralnetworksanddeeplearning.com/

b) Neural Network in Machine Learning- https://www.analyticsvidhya.com/blog/2022/01/introduction-to-neural-networks/

c) Understanding Neural Networks: A Beginner's Guide- https://medium.com/@MakeComputerScienceGreatAgain/understanding-neural-networks-a-beginners-guide-6f719aa88e97

A KING'S GAMBIT FOR A PAUPER'S PRIZE

Rishika Ghosh[1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

"You have 15 Tweets, 7 Retweets and 3 new Followings."

And suddenly, one fine October morning, it all changes to "X" following the acquisition of Twitter by the "self-made" CEO Elon Musk at a deal of US $44 billion (a sum significantly larger than the GDPs of quite many African, Asian and even Central American countries). Following one of the most brutal layoffs in the technology sector - "CEO Parag Agarwal dismissed, more than half of the staff laid off, 'extremely hardcore' work routine ultimatums issued and even more resigned" - read the headline. The user base was quick to follow suit with more than 32 million users lost worldwide by 2024. The platform's value has plummeted by a whopping 71.5%, with X currently valued at $12.5 billion - a mere fraction of the "greater than GDPs" number that, if judiciously used otherwise, could perhaps have brought more direct changes to the world hunger and climate crisis dynamics than merely being a result of a change in user ratios. Simply put, was the King's gambit worth the Pauper's prize?

(As humans are hardwired to be risk averse, the simple answer to that simple question would be no but the thinking behind the action taken is more complex.)

A part of that thought lies in a phenomenon Economists like to call the "Winner's Curse."

An attempt at self-explanation from the term may suggest to mind the fables of Arabian Nights and the hoarded, cursed treasures - the misfortunes that befall the discoverers of the said treasures - or maybe even a 30% GST on the hard-earned prize money of the world (chess) champion; but the answer lies in an even more interesting marketing and psychological phenomenon.

Imagine a scenario where one has to place a bid on a particular item, say a jar full of coins, in an auction-like setting. There are a total of 100 bidders and, through inspection, it is clear that if we plot a probability density function (PDF) of the bids placed, we would have a graph resembling that of a symmetric distribution, with the following features-

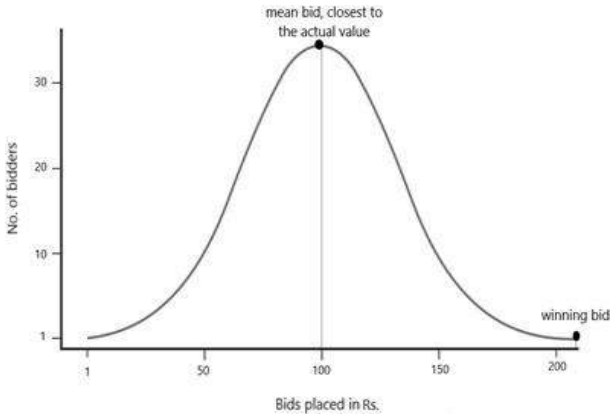


Figure 7.1

- The jar would always have a content of more than Rs.1; hence the minimum possible bet would be the same.
- Now, since most people on average are good at estimating the true value of the jar when full of coins (the dimensions of the jar being known, or estimated coherently), most of the bids would be central around a value, which is close to the true value of the coins contained in the jar. (Regression to the mean)
- Finally, there would be a winning bid (placed by the most enthusiastic bidder, looking at the most optimistic scenario - the jar, able to hold (say) 15 such coins, completely filled with Rs.20 coins) that crosses the mean, or the best-estimated value by a strong margin and sets an overestimated outlier. (An extreme event followed by other, less extreme events)

The curse

It is hence clear that the winner of the auction overestimates or far exceeds the true value of the item. This is the winner's curse, which often translates into the "buyer's remorse" of the most enthusiastic bidder, at having the realization ("curse") of overpayment. Now, when the same

bidding is to be done concerning a company or huge amounts of shares (like that of the acquisition of Twitter), the true value of the shares involved is extremely difficult to generate or predict due to the numerous factors affecting them (Example: future revenues and capitals), unless the bidder is privy to certain inside information that the others are unaware of. Until then, there will always be a margin between the true (mean or estimated) value and the highest (winner's) bid- giving rise to the winner's curse.

Hence, the following conditions give rise to a Winner's Curse situation:

- There are many bidders (by definition, one bidder makes it a sale and not an auction), the exact number may even be unknown in many cases. This makes it harder to correctly estimate the intrinsic value of the prize, often leading to its overestimation. Again, the more interested the parties, the greater the likelihood of an overly enthusiastic bid.
- The many unknowns ensure that no one is really aware of the actual value of the asset, given that the auction is fair and the knowledge about the asset made public, or rather not made public, to the bidders is the same for everyone.
- More often than not, the bidders are not all rational - engaged in a test of will to outdo their competitors (that too, by a good margin); emotional factors like envy, overconfidence, retaliation and biases create the perfect concoction for bad decisions. (*"After all, it is not greed that drives the world, but envy."* – Warren Buffet)

It is also interesting to note that the prize must not have personal sentiment or values attached, for the Winner's curse to apply.

The cursed sectors

Initially coined by Atlantic Richfield Petroleum Engineers - Capen, Clapp and Campbell in 1971 - who observed the poor returns of companies bidding for offshore oil drilling rights in the Gulf of Mexico, and (thus) the consistently winning companies often going down under a few years, the Winner's curse now has applications in any purchase where overbidding is a possibility, including auctions and negotiations.

- In many business acquisitions, the winning bid is largely a result of overestimation of future benefits and succumbing to competitive pressures (Example: The acquisition of Twitter). The same often even drives smaller business decisions like the selection of a project - it is often the most promising, rather the most optimistic project that is chosen (courtesy, the incentive of the respective project managers), which in practice, turns out to be more expensive and time consuming than it was projected to be.
  <u>These decisions are then followed through due to the belief that it would eventually pay off big owing to the Monopoly power and network effects associated with the industries.</u> (Reason why X (Twitter) is still being kept alive even after the huge losses).
- Bull stock markets, characterized by rising prices and increasing demand for the stocks, create an auction-like environment where, as the price of equities climbs more and more, people get enticed into making higher bids and driving the prices even higher. Thus, the people who agree to buy the stocks at inflated prices are overly optimistic- just like the auction's winner.
- In Machine Learning, algorithms are trained to select models with the best performance on a given dataset. This training often sets aside the fact that the chosen model might have been overfitted to the specific training data- leading to optimistic performance estimates; and when applied to new, unseen data, the performance may be significantly worse.
- In experiments with multiple comparisons, researchers may focus on the findings with the most statistically significant results. These "winning" findings may have a larger effect size due to chance - leading to the over-interpretation of their importance.

Last but not the least, it is not necessary for the Winner's Curse to only apply to an "auction-like setting" with many competing options. It would also unsettle me if while shopping for shoes, the seller merrily agreed to my first quoted bargain price, for a pair I thought would have higher value.

Lifting the curse

- In Statistics and related fields, Replication studies and adjusting for multiple comparisons like Bonferroni's correction can help control the false discovery rate while conducting multiple tests and also help confirm or refute initial findings (to reduce the impact of the Winner's Curse).
- For Machine learning, techniques like cross-validation help assess model performance more accurately while reducing overfitting.
- In auction-like settings, it is always imperative to acknowledge the Winner's Curse and try to gather as much information about the asset being auctioned. This is because when one wins an auction, what one often wins is the right to pay more than what everyone else thought the asset to be worth. Thus, psychologically sound people like Warren Buffet often follow a simple rule for the same settings- *"Don't Go!"*

And if one cannot help but be a part of the overbidding saga, there seems to be a second rule coined by the Engineers who conceived the term - "the less information one has compared to his opponents (or the more uncertain the intrinsic value of the asset seems), and the more bidders there are - the lower one should bid".

Although this seems counterintuitive for a person not knowing about the curse, this is what makes an efficient bidder. Efficient bidders rarely win, and the winners are rarely happy. Hence, auctions, much like life, are in the end an act of letting go - the most we can do is make sure our decision is informed.

References

1. Adam Hayes. "Winner's Curse: Definition, How it Works, Causes, and Example". July 21, 2024. https://www.investopedia.com/terms/w/winnerscurse.asp
2. The Decision Lab. "Winner's Curse". https://thedecisionlab.com/reference-guide/psychology/winners-curse
3. Safal Niveshak. "Latticework of Mental Models: Winner's Curse" https://www.safalniveshak.com/latticework-mental-models-winners-curse/

FROM CHANCE TO CONSTANTS: ESTIMATING IRRATIONALS WITH THE HELP OF UNIFORMITY

Rohan Mondal [1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

INTRODUCTION:

In the field of mathematics, irrational numbers such as π and Euler's number(e) play a critical role across various domains, including geometry and calculus. Unlike rational numbers, which may be represented in a form $\frac{p}{q}$ such that p and q are co-prime integers and $q \neq 0$, the irrational numbers possess non-repeating and non-terminating decimal expansions. This distinctive characteristic presents challenges in their estimation. While conventional methods often rely on intricate calculations, elegant statistical techniques can also be employed to approximate these constants.

This article examines the use of uniform random variables to estimate values of π and e. By generating random points within specified ranges, we can utilize techniques such as Monte Carlo simulations to get estimates for these constants. This approach illustrates how randomness can yield unexpectedly accurate approximations, effectively bridging the gap between uncertainty and precision. This exploration not only deepens our understanding of these significant constants but also showcases the efficacy of statistical approaches in the realm of mathematics.

Objective:

This article aims to provide an efficient estimation of the mathematical constants π and e by leveraging classical statistical methodologies, specifically employing key concepts of Large Sample Theory, such as Consistency and Monte Carlo Simulation techniques.

Consistency:

Suppose we have a random variable X with pdf $f_\theta(x)$ where θ is the unknown parameter of the distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n.

Let $T_n = T(X_1, X_2, \ldots, X_n)$ be a sequence of estimator for θ.

Definition of Consistency: A sequence of estimators $\{T_n\}$ is said to be consistent for θ if $T_n \xrightarrow{P} \theta$. i.e., T_n converges in probability to θ. [N.B. we will be using "$\xrightarrow{P}$" symbol to denote convergence in probability.]

Mathematically it can be written as $P[|T_n - \theta| > \varepsilon] \to 0$ as $n \to \infty$ for all $\varepsilon > 0$.

Sufficient Condition of Consistency:

Now we state a set of sufficient conditions for consistency –

Let $\{T_n\}$ be a sequence of estimators of θ. Then if the following conditions hold, we say that $\{T_n\}$ is consistent for θ –

1. $E[T_n] \to \theta$ as $n \to \infty$

2. $Var[T_n] \to 0$ as $n \to \infty$

Invariance Property:

If $\{T_n\}$ is a consistent estimator of θ and f(.) is a continuous function then we can say that $\{f(T_n)\}$ is consistent for $f(\theta)$.

Mathematically, $f(T_n) \xrightarrow{P} f(\theta)$.

Monte Carlo Simulation:

Monte Carlo method constitutes a wide range of computational algorithms that depend on multiple instances of random sampling to generate numerical outcomes. Monte Carlo techniques are frequently applied in both physical and mathematical problems, and they are particularly advantageous when other methods are hard or unfeasible to implement. Basically, by generating a large number of

random inputs, it simulates various scenarios to estimate probabilities and predict outcomes. Here we will compute the value of the estimators for large number of sample sizes and we will get to see that the value of the estimate converging to the value of the expectation of the estimators.

Software Used: For simulating all the procedures and graphs we have used the R software.

Estimating π Using Unit Circles And Uniform Random Variables:

We first need to find a consistent estimator of π. For that define the following setup-

Suppose, we have a random sample of size n such that the sample points are inside a square of side length 2 units centred at the origin.

Now, define a random variable X, which denotes the number of points falling inside an inscribed circle with centre being at the origin and the radius being 1 unit. It is clear that X follows a Binomial Distribution with parameters n and p. Here p is the probability of a point falling inside the circle.

Note that, the probability of a point falling inside the circle is simply the ratio of the area of the circle and the area of the square.

Now here the area of the circle is $\pi*1^2=\pi$ and the area of the square is $2^2=4$.

Dividing the area of the circle by the area of the square we get the ratio of the two areas as:

$$\frac{\text{Area of circle}}{\text{Area of square}} = \frac{\pi}{4}$$

This means that $p = \frac{\pi}{4}$.

Hence $E(X) = \frac{n\pi}{4}$, and $Var(X) = \frac{n\pi}{4}.\frac{3\pi}{4} = n.\frac{3\pi^2}{16}$

Therefore $E(\frac{4X}{n}) = \pi$, also note that $Var(4 * \frac{X}{n}) = \frac{16}{n^2} * Var(X) = 3*\frac{\pi^2}{n}$.

Therefore $Var(4*\frac{X}{n}) \to 0$ as $n \to \infty$.

So, it can be said that by sufficient condition of consistency, $\frac{4X}{n}$ is consistent for π.

Simulation:

Now we will use the Monte Carlo algorithm and the same setup defined earlier. The steps are as follows-

1. We draw a random sample of size n such that the sample observations are inside a square of side of length 2 units centered at the origin.

2. We compute the value of the estimator that we have just defined in the previous part that is, $4 * \frac{X}{n}$.

3. Now we repeat the previous 2 steps for different values of n and we would get different values of our estimator.

After the simulation processes, we summarize the results in the following table and the graph given below.

Table 1. Estimate of π vs sample size

Sample size	10^3	10^4	10^5	10^6	10^7
Estimated value	3.112	3.1416	3.1433	3.1381	3.1416

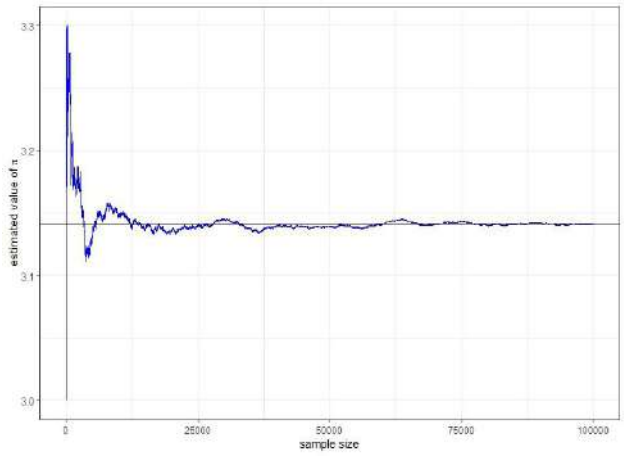


Figure 8.1: Approximation of π

Observation:

Here we can see as the sample size increases our estimate converges to the true value of π.

The above graph (fig.8.1) also shows the rate of the convergence. This also happens to show that the accuracy of our estimator increases as the total number of sample points increases.

Estimating Euler's Number (e) Using Consistency of Mean of Uniform Random Variables:

We first need to find a consistent estimator of the Euler's number e. For that purpose, we are stating and proving the following result.

Result:

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, and each $X_i \sim U(0,1)$. Then sample geometric mean is a consistent for 1/e.

Proof:

- To demonstrate the consistency of the sample geometric mean we start by defining the sample geometric mean of the first n independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots\ldots, X_n$ where each $X_i \sim U(0,1)$.
- The sample geometric mean is given by:
$$\bar{X}_g = \sqrt[n]{\prod_{k=1}^{n} X_k}$$
- Now to simplify calculations we take natural logarithm on both sides we get
$$\ln(\bar{X}_g) = \frac{1}{n} * \sum_{i=1}^{n} \ln(X_i)$$
- In the next step, we need to find the distribution of $-\ln(X_i)$.
- The probability density function of X_i is given by:

$$f(x) = \begin{cases} 1 & \text{, if } 0<x<1 \\ 0, & \text{otherwise} \end{cases}$$

- Let, $F_X()$ denote the C.D.F. of X and given as follows:

$$F_X(x) = \begin{cases} 0, \text{ if } x \le 0 \\ x, \text{ if } x \in (0,1) \\ 1, \text{ if } x \ge 1 \end{cases}$$

- Let, $Y = -\ln(X)$ i.e., $y = -\ln(x) \Rightarrow x = e^{-y}$. Now the range of Y is from $(0, \infty)$. Because if $x=0$ then $y \to \infty$ and if $x=1$ then $y=0$.
- Let $G_Y()$ be the C.D.F. of Y.

  Now, $G_Y(y) = P[Y \le y] = P[-\ln(X) \le y] = P[X \ge e^{-y}] = 1 - P[X \le e^{-y}] = 1 - e^{-y}$.
- Therefore the C.D.F. of Y is given by,

$$G_Y(y) = \begin{cases} 0 & \text{, if } y<0 \\ 1-e^{-y} & \text{, if } y \ge 0 \end{cases}$$

- Now from the uniqueness property of C.D.F. of a random variable, we can say that $Y \sim exp(1)$. That is Y follows an exponential distribution with mean 1. That means $E[Y_i] = 1$ and $Var[Y_i] = 1$.
- Here note that as X_i's are independent so Y_i's, being a function of X_i's, will also be independent.
- We define $S_n = \sum_{i=1}^{n} Y_i$
- Now $E(S_n) = \sum_{i=1}^{n} E(Y_i) = n$. and $Var(S_n) = \sum_{i=1}^{n} Var(Y_i) = n$. [all the covariance terms are 0 as Y_i's are independent]. Hence $E(\frac{S_n}{n}) = 1$ and $Var(\frac{S_n}{n}) = \frac{1}{n}$.
- Now $Var(\frac{S_n}{n}) \to 0$ as $n \to \infty$. Therefore, by the sufficient conditions of consistency we can say that $\frac{S_n}{n} \xrightarrow{P} 1$.
- We can also say that $-\frac{S_n}{n} \xrightarrow{P} -1$. [using the properties of convergence in probability]
- $\ln(\bar{X}_g) = \frac{1}{n} * \sum_{i=1}^{n} \ln(X_i) = -\frac{1}{n} * \sum_{i=1}^{n} -\ln(X_i) = -\frac{1}{n} * \sum_{i=1}^{n} Y_i = -\frac{S_n}{n}$
- Therefore $\ln(\bar{X}_g) \xrightarrow{P} -1 \Rightarrow \bar{X}_g \xrightarrow{P} \frac{1}{e}$.[using the invariance property of consistent estimators]

So, it means that we have proved our desired result.

Therefore, we can say that $\frac{1}{\bar{X}_g} \xrightarrow{P} e$. [again using the invariance property of consistent estimators].

Simulation:

Now for the Monte-Carlo Method we follow the following steps:

1. We draw a random sample of size n from U(0,1) distribution.

2. We compute the value of the estimator that we have just defined in the previous part which is the reciprocal of the sample geometric mean.

3. Now we repeat the previous 2 steps for different values of n and we would get different values of our estimator.

After the simulation processes, we summarize the results in the following table and the graph given below.

Table 2: Estimate of e vs sample size

Sample size	10^3	10^4	10^5	10^6	10^7
Estimated value	2.7513	2.7513	2.7202	2.7180	2.7182

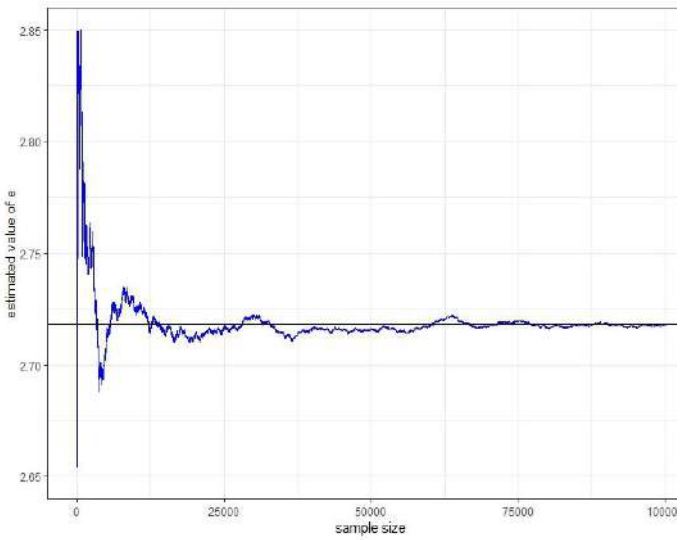


Figure 8.2: Approximation of e

Observation:

Here we can see as the sample size increases our estimate converges to the true value of e. The above graph (fig. 8.2) also shows the rate of the convergence. This also happens to show that the accuracy of our estimator increases as the sample size increases.

Conclusion:

To summarize, we have investigated an intriguing use of uniform random variables to estimate two of the most essential mathematical constants, π and e. By utilizing the concepts of probability and statistical simulation, we have shown how straightforward random experiments can produce surprisingly accurate approximations. This method emphasizes the effectiveness of computational techniques and establishes a tangible link between abstract mathematical ideas and practical simulations. As technology evolves, we anticipate the emergence of even more advanced methods that will enhance our comprehension of these perplexing numbers.

References:

1. S.C. Gupta and V.K. Kapoor, “Fundamentals Of Mathematical Statistics”.

MAKING COMMENT ON CONSISTENCY OF A BATSMAN IN CRICKET USING DISPERSION MEASURE: A VIEW ON THE PERFORMANCE OF INDIAN BATTERS IN 2023 ODI WORLD CUP

Saptaswa Sengupta [1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

Introduction: In our daily lives, it is very much important to analyze the consistency of the given data in our hands. In sports, it is also important as we often want to know how consistent a player is in a particular tournament. When it comes to cricket, our interest lies in how consistently a batsman/batter is scoring runs or how consistently a bowler is taking wickets. In this discussion, we will consider only the batters' performance.

Role of Dispersion in Finding Consistency: Dispersion is a characteristic indicating the extent to which observations vary among themselves. A measure is designed to state numerically the extent to which individual observations vary on the average. There are mainly two kinds of measures of dispersion - Absolute Measure and Relative Measure. Range, Quartile Deviation, Mean Deviation, and Standard Deviation are considered Absolute Measures. On the other hand, Coefficient of Variation, Coefficient of Mean Deviation, and Coefficient of Quartile Deviation are considered Relative Measures.

In most cases, we use Standard Deviation and Coefficient of Variation (C.V.) in analyzing the consistency of the given data sets because the other measures are not as good as these measures. Range does not depend on all observations, Quartile Deviation does not take into account the variability of all observations and in the case of Mean Deviation, the procedure of neglecting the signs and taking absolute deviations only makes algebraic treatment difficult.

If the average values of the datasets are markedly different, then to measure the variability of those datasets we should use relative measure instead of absolute measure. Less relative measure/absolute measure reflects more consistency. Similarly, more relative measure/absolute measure reflects less consistency.

Who was the most consistent batsman for Team India in the 2023 ICC Cricket World Cup?

In finding out the most consistent batsman for team India in the recent ODI World Cup, we feel to include only the specialist batters' and all-rounders' batting performances.

An important note:

- Here, we do not include the bowlers' batting performances because their batting performances are not so remarkable.
- We will consider the batting performances of only those batters who did bat in at least 7 innings (not matches). Otherwise, we will face problems in comparing their performance because the batters, who batted in the top order, got the chance to bat in almost every match whereas some of the middle-order and the lower-order batters did not get so many chances to bat. So, if we consider the case of all batters, it may lead us to a biased conclusion.

In this World Cup, India has played a total of 11 matches including the Semi-Final and the Final. In those 11 matches, a total of 9 batters have batted (7 specialist batters and 2 all-rounders).

Specialist batters' names: Rohit Sharma (Captain), Shubhman Gill, Virat Kohli, Shreyas Iyer, KL Rahul (Wicket Keeper-Batsman),

Suryakumar Yadav, Ishan Kishan. All-rounders' names: Ravindra Jadeja, Hardik Pandya.

It is to be noted that in this World Cup, Ishan Kishan played only 2 matches, Hardik Pandya played only 4 matches where he batted only in the first match (vs Australia), Ravindra Jadeja played all the matches but batted in only 5 matches and Suryakumar Yadav batted in less than 7 matches. So, as per our consideration, we will work with the batting performance of the other 5 batters: Rohit Sharma, Shubhman Gill, Virat Kohli, Shreyas Iyer and KL Rahul.

The following table shows the number of runs scored by each of those six batsmen at each match:

Batters	Rohit Sharma	Shubman Gill	Virat Kohli	Shreyas Iyer	KL Rahul
vs Australia	0	0	85	0	97
vs Afghanistan	131	0	55	25	0
vs Pakistan	86	16	16	53	19
vs Bangladesh	48	53	103	19	34
vs New Zealand	46	26	95	33	27
vs England	87	9	0	4	39
vs Sri Lanka	4	92	88	82	21
vs South Africa	40	23	101	77	8
vs Netherlands	61	51	51	128	102
vs New Zealand (Semi-Final)	47	80	117	105	39
vs Australia (Final)	47	4	54	4	66

Table 9.1

Now, supposed to comment on the consistency of these five batters, we decide to proceed with Standard Deviation (S.D.). The Standard Deviations (S.D.) of five batters are given below:

Rohit Sharma's S.D.=37.28

Shubhman Gill's S.D.=32.31

Virat Kohli's S.D.=37.53

Shreyas Iyer's S.D.=44.09

KL Rahul's S.D.=33.75

A lower S.D. indicates that the batter's scores are more consistent, with fewer extreme variations. Thus using Standard Deviation (S.D.), we have come to the conclusion that Shubhman Gill is the most consistent batter since he has the lowest S.D. among all batters and Shreyas Iyer is the most inconsistent batter as he has the highest S.D. among all batters. But, in this case, there is a major problem while we are using S.D. as a measure of consistency. To understand the problem, first, we will calculate the average of each of these five batters, i.e., we will calculate their means. It is to be noted that we will calculate their means on the basis of (total runs/total innings) instead of (total runs/total matches).

Thus, we get:

Average of Rohit Sharma = 54.27

Average of Shubhman Gill = 32.18

Average of Virat Kohli = 69.54

Average of Shreyas Iyer = 48.18

Average of KL Rahul = 41.09

Note that the averages of five batters are markedly different. Thus, there will arise some problems if we comment on consistency only on the basis of S.D..

S.D. is the measure of only spread or

variability but it doesn't take into account the average performance. A batter with high S.D. and high average indicates that the batter scores runs in burst but also has a period of low scores while a batter with lower S.D. and lower average indicates that the batter consistently scores lower runs.

For example, we can consider the case of Shreyas Iyer and KL Rahul where Shreyas has a higher average and higher S.D. than Rahul. Now if look only at S.D., we will see that S.D. of Virat Kohli (37.53) and S.D. of Shubhman Gill(32.31) are very close. But if we look at their averages, we will see that the average of Virat Kohli (69.54) is much more than the average of Shubhman Gill (32.18). Now, that batsman is more desirable for us, who has a high average and a low S.D., i.e. the batsman who scores runs regularly with minimum variation in runs. Hence, while we looked at only S.D. we had concluded that the most consistent batter (Gill) has the lowest S.D. and the lowest average among all batters, which is not desirable.

Thus, using both average and S.D. provides a more comprehensive picture of a batter's consistency. Coefficient of Variation (C.V.) serves that purpose as it is a measure based on both mean and S.D. Thus, we will use C.V. as a measure of consistency in this case.

Now, Calculating the C.V.s of each of the five batters, we get:

Rohit Sharma's Coefficient of Variation (CV1) = 68.70%

Shubhman Gill's Coefficient of Variation (CV2) = 100.40%

Virat Kohli's Coefficient of Variation (CV3) = 53.97%

Shreyas Iyer's Coefficient of Variation (CV4) = 91.50%

KL Rahul's Coefficient of Variation (CV5) = 82.13%

Conclusion:

Thus, it is clearly observed that among these five batters, Virat Kohli has the lowest C.V. (53.97%) and Shubhman Gill has the highest C.V. (100.40%). As higher C.V. implies lower consistency and lower C.V. implies higher consistency, therefore, Virat Kohli is considered to be the most consistent batsman for Team India in 2023 Cricket World Cup. On the other hand, Shubhman Gill is considered to be the most inconsistent among these five batters of Team India. Also, Rohit Sharma (second most consistent batter), Shreyas Iyer, and KL Rahul consistently scored good runs but they were not as consistent as Virat Kohli. India's most consistent batsman Virat Kohli was the key performer in the World Cup as he was the highest run scorer of that World Cup (765 runs) and he also achieved the prize of the "Player of the Tournament". Though team India failed to fulfil the expectation of winning the World Cup as they lost in the Final after 10 consecutive wins, this World Cup will surely be memorable for their stunning performances in batting, bowling, and fielding.

References:

1) Fundamental Statistics (Vol – I): AM Gun, MK Gupta, B Dasgupta

2) Introduction to Statistics: Prasanta Kumar Giri, Jiban Banerjee

3) Statistical Methods: NG Das

4) Anandabazar Patrika

5) https://www.espncricinfo.com

COMBATING THE INVISIBLE

Shreshtha Sengupta [1]

M.Sc Data Science, St. Xavier's College (Autonomous), Kolkata

India, the land which has been treasuring the Eternal Truth of the Vedas since time immemorial, is now advancing into the age of "Digital Revolution" with all her glory and power. Thus, be it the Edicts of the 3rd-century Emperor Ashoka in Kalinga or be it the Covid-19 vaccine details of your kith and kin in the CoWIN portal, proper documentation of data has always been considered a 'de rigueur' in the administrative processes of this nation, since days of yore.

According to the data from MyGovIndia, with 89.5 million digital transactions in the year 2022, India has topped the list of five countries in digital payments in the whole world. There is now a total of 750 million active Internet users in the country. Up to the FY 2022, the number of recognized startups has increased to a whopping 80,152. The statistics mentioned should be enough to signify the vastness of the *'ocean of data'* that is being shared, stored, and crunched every second in India, let alone the whole world.

A concerned citizen of yesterday might have only taken interest perhaps in knowing about the local law & order situation of his/her own locality, but the inquisitive netizen of today takes an avid interest in world geopolitics and the way his/her own country contributes to the welfare of the world. 'Assimilation not Destruction!' was Swami Vivekananda's message to the world, and India stands as an epitome of harmony and tolerance to this day. But unfortunately, India has always been at the evil snares of bloodthirsty devils since ages, and thus, when it comes to the defence of her borders from enemies, India now produces the best of soldiers and is constantly making efforts in making improvements in her defence infrastructure and equipping herself with modern ammunition.

Does a skilful enemy always attack from the front? No, never! Netizens may not realise this now, but there are numerous security threats which are staring at them, right through their shining smartphone screens, as merciless, 'data-thirsty' demons!

Be it hospitals, bank payments or online shopping, each one of us is tremendously dependent on technology for our day-to-day necessities. Gone are the days when one used to take huge pains of filling out college admission forms and writing paper bills manually. But, by mindlessly sharing sensitive information over the internet, both knowingly and unknowingly, are we putting ourselves in grim danger? Do the social media websites simply hoodwink us into believing that our photographs are completely safe with them and our private chats are 100% encrypted?

Well, in the year 2017, The Economic Times, a leading business newspaper of India, approached several 'data-brokerage' companies which apparently sell sensitive information of both working and retired professionals at a very minimal cost to literally anyone on the dark web. The ET gathered from a broker that a database of 1.7 lakh people from Delhi, NCR and Bengaluru, which included credit card numbers of young working professionals, purchase details of customers from online shopping portals like Amazon and phone numbers of senior citizens, can be made available for just Rs 7,000!

The dark horses behind such crimes are quite adept at using the power of deception to infiltrate organisations around the globe with malware by persuading the employees into providing information or performing an action that would benefit the attacker. This is known as "***Social Engineering,"*** a term coined by the world-famous hacker Kevin Mitnick in his

book "The Art of Deception." Kevin says in his book- "most people go on the assumption that they will not be deceived by others, based upon a belief that the probability of being deceived is very low; the attacker, understanding this common belief, makes his request sound so reasonable that it raises no suspicion, all the while exploiting the victim's trust."

Some major types of social engineering are-

1. **Phishing** is a cybercrime in which a target is contacted by email, telephone, or text message by someone posing as a legitimate institution to lure individuals into providing sensitive data.

2. **Baiting** puts something enticing in front of the victim to lure them into the social engineering trap. A baiting scheme could offer a free music download or gift card to trick the user into providing financial credentials.

3. **Pretexting** is often used against corporations that retain client data, such as banks, credit card companies, utilities, and the transportation industry." During pretexting, the threat actor will often impersonate a client or a high-level employee of the targeted organization.

4. **Whaling attack** is a type of social engineering attack specifically targeting senior or C-level executive employees with the purpose of stealing money or information or gaining access to the person's computer to execute further cyberattacks.

In November 2022, the online hospital management system of the prestigious AIIMS, New Delhi, was caught in the clutches of a terrible cyber-attack, which reportedly might have led to the leak of sensitive medical data of around 4 crore patients, including top Government officials, retired defence personnel and, most importantly the citizens.

jointly investigating the case.

The fact that the news of such a large-scale data breach failed to come into the limelight of several leading TV news channels in the country is quite shocking. Not only does this data disaster reveal the lackadaisical attitude of the database administrators of the most reputed medical institution, but also points out the sheer ignorance of the millennials who are not yet taking this threat of cybersecurity to be real.

According to a Singapore-based cybersecurity firm, the cyberattack at the Kudankulam Nuclear Power Plant in 2019 might have remained undetected for six months! This nuclear power plant acts as one of the primary sources of electricity for three major Southern Indian states: Tamil Nadu, Kerala, and Karnataka. A slight mishap in the system of such a large nuclear plant can not only lead to power loss in almost the whole of South India, but could also be a reason for one of the deadliest catastrophes of the world, leading to a loss of lakhs of precious human lives. It still sends a shiver down the spine to even hear about the brutal atomic bombings of Hiroshima-Nagasaki and the infamous Bhopal Gas Tragedy.

Although it is quite heartening to see the setting up of the Defence Cyber Agency, a tri-service command of the Indian Armed Forces for combating cyber-security threats, the delayed security enhancements in the database systems of both government and private owned institutions remains a matter of grave concern. The Digital Personal Data Protection Bill, 2022 is a promising step in the direction of National Data Security, which keeps the protection of personal data as its topmost priority, especially that of children. According to the bill, any data fiduciary which shall undertake any kind of tracking or behavioural monitoring of children will be obliged to pay a high penalty. Moreover, the

bill provides protection to the consumers against important cybercrimes like identity theft, stealing of bank details and transfer of private data by data fiduciaries to third parties without the consent of the consumer. A revised version of the bill was introduced in 2023 and was passed in both the houses of the parliament as the ***Digital Personal Data Protection Act, 2023.***

The "**Cyber Swachhta Kendra**" or the Botnet Cleaning and Malware Analysis Centre is a part of the Government of India's Digital India initiative under the Ministry of Electronics and Information Technology to create a secure cyber space by detecting botnet infections in the country and to notify, enable cleaning and securing systems of end users so as to prevent further infections. This Centre was set up in accordance with the objectives of the "National Cyber Security Policy", which envisages creating a secure cyber ecosystem for India.

Right from the Prime Minister to a street hawker, each citizen of the nation can play an integral part in national cyber security. All it requires is a little effort. Young and tech-oriented adults must pitch in and spread awareness amongst the old, who, owing to their naivety, fall prey to criminals posing as bank officials over the phone, only to steal their hard-earned money. Students can also guide their house-help and newspaper hawkers to tackle the problem of phishing and help them to set up strong passwords because weak, repetitive passwords for social media accounts often pave the way for hackers to steal data. Awareness campaigns can be organised regarding the harmful effects of over-sharing of personal photographs and live locations on social media.

As Rumi said, "Yesterday I was clever, so I wanted to change the world. Today, I am wise, so I am changing myself!" Thus, each one of us can protect the privacy of our dearest nation, along with our soldiers, by being a *"data-literate"* netizen and witness India flying like a strong Golden Bird with infinite strength!

References:

1. https://blog.mygov.in/editorial/the-digital-india-transformation/
2. https://economictimes.indiatimes.com/tech/internet/how-data-brokers-are-selling-all-your-personal-info-for-less-than-a-rupee-to-whoever-wants-it/articleshow/57382192.cms?from=mdr
3. The Digital Personal Data Protection Bill,2022
4. https://cybersrcc.com/2022/12/12/cyber-attack-on-aiims-indian-healthcare/
5. https://timesofindia.indiatimes.com/blogs/ChanakyaCode/cyber-attack-on-kudankulum-nuclear-power-plant-underlines-the-need-for-cyber-deterrent-strategy/
6. Mitnick, Kevin D., and William L. Simon. *The Art of Deception: Controlling the Human Element of Security*. Wiley, 2002.
7. https://www.mitnicksecurity.com/blog/6-types-of-social-engineering-attacks
8. The Digital Personal Data Protection Act, 2023- https://prsindia.org/billtrack/digital-personal-data-protection-bill-2023

GRADIENT BOOSTING MACHINES

Soumyadeep Roy[1]

M.Sc Data Science, St. Xavier's College (Autonomous), Kolkata

Introduction

Gradient Boosting Machines (GBMs) are powerful ensemble learning techniques that iteratively refine predictions by combining multiple weak learners, typically decision trees, into a highly accurate and robust model. By sequentially constructing each new model to correct the residual errors of its predecessors, GBMs effectively minimize bias and variance, leading to superior predictive performance.

Renowned for their dominance in machine learning competitions and real-world applications, GBMs excel in both classification (e.g., spam detection) and regression (e.g., house price prediction). Their ability to strike a balance between predictive power and overfitting control makes them indispensable across diverse domains such as finance, healthcare, and marketing. From fraud detection and risk assessment to patient diagnosis and customer segmentation, GBMs drive data-driven decision-making with precision and adaptability, making them a cornerstone of modern predictive analytics.

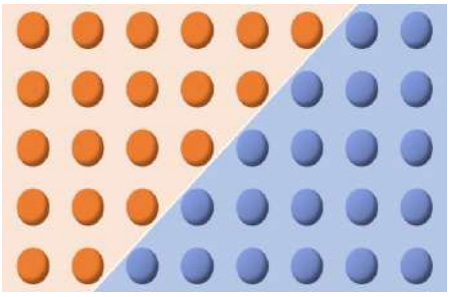


Fig 11.1

Boosting Concept and Gradient Descent

Understanding Boosting

Boosting is an ensemble learning technique that enhances predictive accuracy by sequentially training models, with each new model focusing on correcting the errors of its predecessors. The core idea is to build a strong learner by combining multiple weak learners, typically decision trees, to progressively refine predictions.

A simple analogy is guessing a number: each incorrect guess provides valuable feedback that helps refine subsequent attempts. Similarly, in GBMs, each model learns from the mistakes of the previous ones, iteratively improving performance. By assigning higher weights to misclassified or high-error instances, boosting ensures that each iteration focuses more on the challenging cases, leading to a more robust final model.

Gradient Descent in GBM

The term *Gradient* in Gradient Boosting originates from *gradient descent*, a powerful optimization technique used to minimize errors. In GBMs, gradient descent helps the model to iteratively adjust its parameters to minimize a predefined loss function, ensuring improved predictions over time.

A useful analogy is descending a mountain: you aim to reach the lowest point (optimal solution). By observing the steepness of the slope (gradient), you determine the best direction to move. Taking small, calculated steps downward ensures steady progress toward minimizing errors without overshooting the goal.

Mathematically, Gradient Boosting minimizes the following objective function:

$$L_i = \sum_{i=1}^{n} \mathrm{L}(y_i, \hat{y_i})$$

- L is the overall loss function,
- y_i represents the true values,
- $\hat{y_i}$ represents the predicted values,
- $\mathrm{L}(y_i, \hat{y_i})$ is the loss function measuring the difference between true and predicted values.

Gradient Boosting Algorithm with Example

Building a Gradient Boosting Classification Model

Gradient Boosting iteratively combines multiple weak learners, typically decision trees, to construct a strong predictive model. Each new model is trained to correct the errors of its predecessors, gradually improving overall accuracy.

Consider a simple classification dataset where the target variable has binary class labels (0 and 1) and two features, x_1 and x_2 . The model sequentially learns patterns in the data, refining predictions at each step to enhance classification performance.

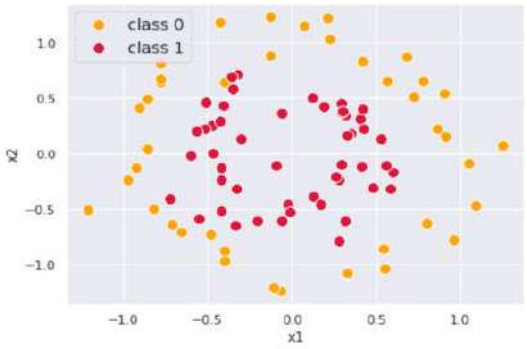


Fig 11.2

Explanation: The plot shows two classes (orange and red) in a circular pattern, indicating non-linearly separable data.

Our goal is to build a gradient boosting model that classifies those two classes.

Building a Gradient Boosting Model for Classification

Step 1: Initialize with a Constant Prediction

The first step in Gradient Boosting is to create an initial constant prediction, denoted as F_0 .Since we are dealing with a classification problem, we use **log loss** (also known as cross-entropy loss) as our loss function. The initial *prediction is chosen to minimize this loss:*

$$F_0 = argymin_i = \sum_{i=1}^{n} \mathrm{L}(y_i, \hat{y}_i)$$

are dealing with a classification problem, we use **log loss** (also known as cross-entropy loss) as our loss function. The initial *prediction is chosen to minimize this loss:*

$$F_0 = argymin_i = \sum_{i=1}^{n} \mathrm{L}(y_i, \hat{y}_i)$$

- $\mathrm{L}(y_i, \hat{y}_i)$ is the loss function (log loss in this case),
- y_i represents the actual class labels,
- γ is the initial prediction value.

Since we are performing binary classification (with class labels 0 and 1), a reasonable initial prediction is the proportion of class 1 in the dataset. This is computed as:

$$P = \sum_{1}^{n} yi \,/\, n$$

where P represents the probability of class 1 across all data points. This ensures that the initial model starts with the best possible uniform prediction before refining it in subsequent steps.

$$p = P(y = 1) = \bar{y}$$

Here is a 3D representation of the data and the initial prediction. At this moment, the prediction is just a plane that has the uniform value $p = mean(y)$ on the y axis all the time.

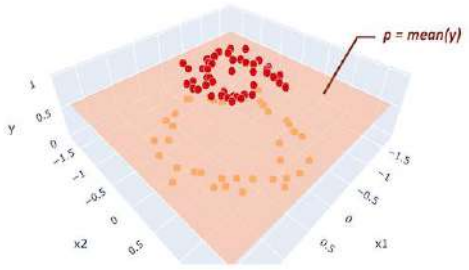


Fig 11.3

Explanation: This 3D plot visualizes two non-linearly separable classes, with a decision boundary at p = mean(y) highlighting class separation.

Step 2: Understanding Residuals in Gradient Boosting

The loss function $\mathrm{L}(y_i, \hat{y}_i)$ measures the discrepancy between the true values y and the predicted values $\hat{y}_i$. In classification tasks, we

commonly use **log loss**, while for regression, **mean squared error (MSE)** is a standard choice.

Components of the Loss Function:

- L represents the loss function.
- y represents the true class labels.
- $\widehat{y_l}$ represents the predicted values (initially uniform).
- In our dataset, the mean of y is **0.56**. Since this value is greater than **0.5**, our initial model classifies all observations as class **1**. While this uniform prediction might seem overly simplistic, it serves as a baseline that will be refined with additional weak models.

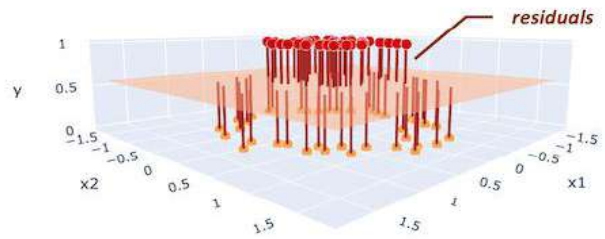


Fig 11.4

Explanation: This 3D plot visualizes residuals, showing the differences between actual and predicted values, with vertical lines indicating errors from the decision boundary.

Step 4: Fitting a Regression Tree to Residuals

To minimize the residuals and improve our model's predictions, we now build a **regression tree** using the features $x_1$andx_2, with the residuals r as the target variable. The goal of this tree is to identify patterns in the relationship between the features and residuals, allowing us to systematically reduce errors from the initial prediction p.

Why Use a Regression Tree?

- The tree learns how different feature values influence residuals.
- By finding structure in the residuals, it helps adjust predictions in the right direction.
- The tree's output provides corrections that, when added to p, refine the overall model.

Creating the First Regression Tree

We fit a decision tree to predict the residuals and obtain **two distinct leaf values**, where:

r = {0.1, -0.6}

This means the tree splits the dataset into two regions based on the values of x_1 and x_2, assigning a predicted residual of **0.1** to one region and **-0.6** to the other. These values represent the model's best guess for adjusting the initial prediction in each region.

At this stage, our model is still relatively simple, but as we iteratively add more trees, it becomes increasingly refined, gradually reducing errors and improving classification accuracy.

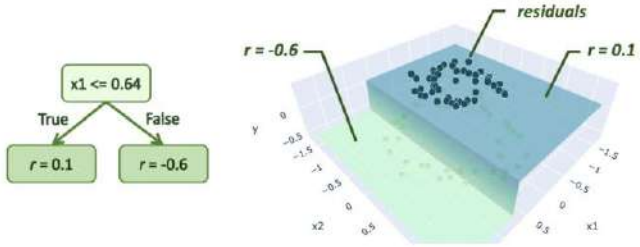


Fig 11.5

Explanation: This image illustrates a decision tree split (left) based on x1, assigning different residual values to two regions. The 3D plot (right) visualizes these residuals, showing the decision boundary and error distribution.

The values (we call it γ gamma) that we are adding to our initial prediction is computed in the following formula:

$$\gamma_j = \frac{\left\{\sum_{\{x_i \in R_j\}}(y_i - p)\right\}}{\left\{\sum_{\{x_i \in R_j\}} p(1 - p)\right\}}$$

Step 5: Updating Predictions Using Log-Odds and Learning Rate

Understanding Aggregation in Terminal Nodes

In Gradient Boosting, after fitting a regression tree to the residuals, we compute a **terminal node-specific adjustment** γj for each leaf node. This

adjustment is computed as:

$$\gamma_j = \frac{\left\{\sum_{\{x_i \in R_j\}} (y_i - 0.56)\right\}}{\left\{\sum_{\{x_i \in R_j\}} 0.56(1 - 0.56)\right\}}$$

For our two terminal nodes:

$$y_1 = 0.3, \qquad y_2 = -2.2$$

$$\gamma_1 = \frac{\left\{\sum_{\{x_i \in R_1\}} (y_i - 0.56)\right\}}{\left\{\sum_{\{x_i \in R_1\}} 0.56(1 - 0.56)\right\}} = 0.3$$

$$\gamma_2 = \frac{\left\{\sum_{\{x_i \in R_2\}} (y_i - 0.56)\right\}}{\left\{\sum_{\{x_i \in R_2\}} 0.56(1 - 0.56)\right\}} = -2.2$$

Converting Predictions to Log-Odds

Rather than directly adding γ to our initial probability prediction p=0.56, we first convert p into **log-odds** using the transformation:

$$\log(odds) = \log_e \frac{p}{(1-p)}$$

Step 6: Scaling γ with Learning Rate ν to Prevent Overfitting

To ensure that our model does not **over fit** to the training data, we introduce a **learning rate** ν (also called a **shrinkage parameter**). This learning rate scales down γ before updating the model.

*Why Scale **γ** with **ν**?*

- Without scaling, the model might overcorrect and fit noise in the training data.
- A lower learning rate helps the model generalize better by making smaller, incremental updates.
- It ensures a **gradual optimization** process rather than drastic adjustments.

Applying Learning Rate to the Log-Odds Prediction

Instead of directly updating the model as:

$F(x) = F_0 + \gamma$

We modify it using the learning rate $\nu \backslash nu\nu$:

$F(x) = F_0 + \nu . \gamma$

$$F_1(x) = F_0(x) + \nu \cdot \gamma$$

Updated prediction — Initial prediction — Learning rate

*Choosing the Learning Rate **ν** and Converting Log-Odds Back to Probability*

For demonstration purposes, we use a **relatively large learning rate** ν=0.9 to make the optimization process easier to follow. However, in practical applications, **a much smaller value (typically ν=0.1 or lower)** is preferred to ensure **better generalization** and prevent overfitting.

*Converting Log-Odds **F(x** to Probability **p(x))***

After updating the log-odds prediction:

$F(x) = F_0 + \nu \cdot \gamma S$

we convert it back into probability using the **sigmoid function**:

$$p(x) = 1/\{1 + e^{-F(x)}\}$$

Effect on Probability Predictions

- Since the updates to F(x) happen in discrete steps (after adding weak learners), the probability function p(x) forms a **stair-like structure** rather than a smooth curve.
- With more boosting iterations, the probability estimates become **more refined**, reducing misclassifications.

This stepwise correction process is what makes **Gradient Boosting Machines (GBMs) powerful**, as they iteratively improve weak predictions while maintaining control over model complexity.

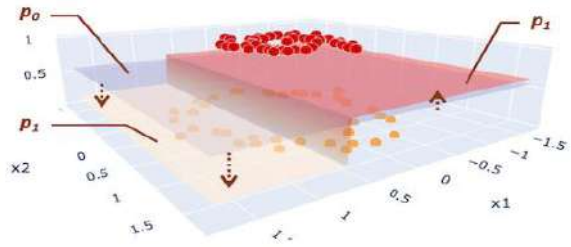


Fig 11.6

Explanation: This 3D plot visualizes class probabilities (p0 and p1) for a classification model. The plane represents the decision boundary, separating two classes with different probability regions.

The purple-coloured plane is the initial prediction p0 and it is updated to the red and yellow plane p1.

Now, the updated residuals r looks like this

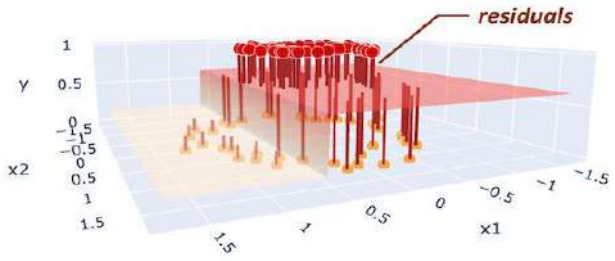


Fig 11.7

Explanation: This 3D plot visualizes the residuals, representing the difference between predicted and actual values. The vertical lines indicate the magnitude of these errors, with larger lines signifying higher discrepancies.

We are creating a regression tree again using the same x_1 and x_2 as the features and the updated residuals r as its target.

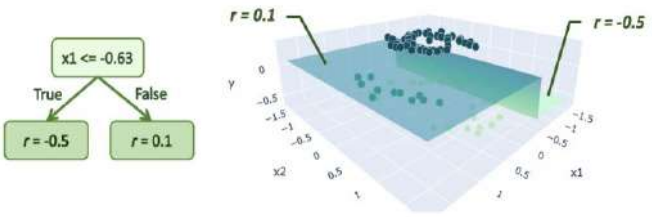


Fig 11.8

Explanation: This image represents a decision tree split on $x1 \leq -0.63$, with two regions assigned different residual values.

The 3D plot visualizes these residuals, showing how the tree partitions the space and assigns predictions.

$F_2(x) = \{F_1(x) - v*2.3 = 0.5 - 0.9*2.3 = -1.6$ if $x_1 \leq -0.63$
$F_1(x) + v*0.4 = 0.5 + 0.9*0.4 = 0.9$ else if $-0.63 < x_1 \leq 0.64$
$F_1(x) + v*0.4 = -1.7 + 0.9 * 0.4 = -1.3$ otherwise.$\}$

These are y computed with this formula:

$$\gamma_j = \frac{\left\{\sum_{\{x_i \in R_j\}} (y_i - p)\right\}}{\left\{\sum_{\{x_i \in R_j\}} p(1-p)\right\}}$$

If we convert log-odds $F_2(x)$ back into the predicted probability $p_2(x)$, it looks like something below:

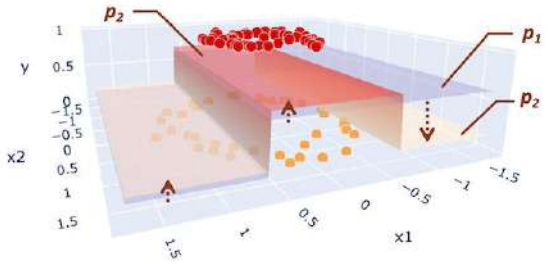


Fig 11.9

Explanation: This 3D plot shows a **decision boundary** in a classification model, likely from a **decision tree**. The space is divided into **three regions**, each assigned a probability (p_1, p_2).

- **Red and orange dots** represent data points from different classes.
- **Planes separate the regions**, showing the model's decision-making.
- **Arrows highlight probability values** assigned to each region.

This visualization helps understand how the model classifies data in different feature spaces.

We iterate these steps until the model prediction stops improving.

When to Stop Gradient Boosting?
Stopping criteria in GBMs includes:

1. **Early Stopping** – Monitor validation loss and stop when it stops decreasing.
2. **Fixed Number of Iterations** – Set a predefined limit (e.g., 100, 500, 1000 iterations).
3. **Performance Threshold** – Stop when improvement is below a set threshold.

4. **Monitoring Residuals** – Stop when residuals are small and randomly distributed.
5. **Gradient Magnitude Threshold** – Stop when average gradient approaches zero.

Best Practice:

Using **early stopping** with a validation set ensures the model generalizes well without overfitting.

1. Convergence of Loss Function

- The model stops when the loss function $L(y, \hat{y})$ no longer significantly decreases.
- Mathematically, if the change in loss between consecutive iterations is below a threshold:
- $| Lt - Lt - 1 | < \epsilon |$ where Lt is the loss at iteration t, and ϵ is a small constant $(e.g., 10e4)$.

2. Early Stopping with Validation Data

- Use a separate validation set to monitor generalization performance.
- Stop when the validation loss starts increasing: $Lval(t) > Lval(t - k)$ for some patience parameter$k$, meaning performance has not improved over k iterations.

3. Gradient Magnitude Threshold

- The algorithm stops if the gradient updates become too small, indicating convergence.
- If the average gradient norm $| \nabla L |$ is below a threshold δ.
- $(1/n) \sum_{i=1}^{n} \nabla L(y_i, \hat{y_i}) < \delta$ Where n is the number of data points.

The image below summarizes the whole process of the algorithm:

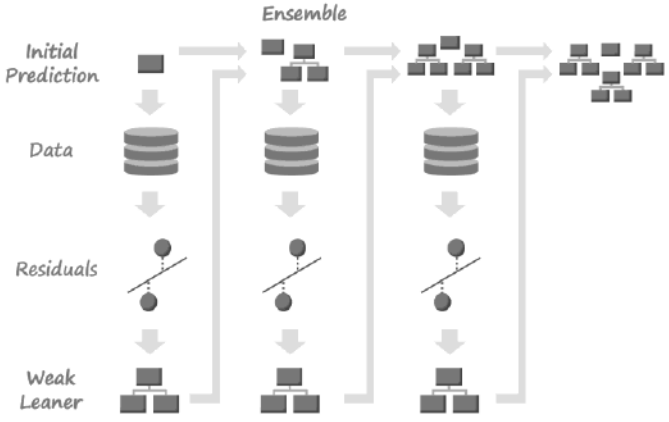


Fig 11.10

Therefore, the algorithm is given by:

Gradient Boosting Algorithm

1. Initialize model with a constant value:

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma)$$

2. for $m = 1$ *to* M:

2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ *for* $i = 1,\ldots, n$

2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\ldots, J_m$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$ *for* $j = 1,\ldots, J_m$

2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

References:

1.https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-2-classification-d3ed8f56541e/

2.https://www.youtube.com/watch?v=4p5EQtyxSyI

3. https://www.geeksforgeeks.org/ml-gradient-boosting/

4. https://www.linkedin.com/pulse/gradient-boosted-algorithm-explained-damien-benveniste-qci3c/

THE NEWSVENDOR MODEL

Srija Upadhyay[1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

Problem: The newsvendor model is a mathematical model that helps determine the ideal inventory level for a product with uncertain demand and a limited shelf life. It can be imagined as the problem of deciding the number of newspapers a newsvendor shall buy to meet his demand of newspapers, such that his profit is maximised. A newspaper has no value if stored, it has to be sold on the same day as it is printed. Also, the vendor can buy the newspapers at once in the morning only, i.e., there is no scoping of buying after realising the day's demand. Thus, the measurement of inventory level is critical to be determined.

Stepping into the problem: Say, a newsvendor buys newspapers for Rs c each and sells at Rs p each, where p>c. So, he makes a profit of Rs (p-c) on selling each newspaper and makes a loss of Rs c for not being able to sell one newspaper. Also, he loses Rs (p-c) if he cannot meet the demand of each extra newspaper than his stock.

So, the vendor faces two types of cost:

a) Cost of Underage: Opportunity cost of turning away a customer.
b) Cost of Overage: Cost of discarding an unsold newspaper.

Our motive is to find the profit maximizing quantity of newspaper that the newsvendor should buy.

Let D denote a random variable denoting the demand of newspaper on a randomly selected day.

F(.) denotes the CDF of D. Let p denote the selling price of one newspaper and c denote the cost price of one newspaper. Also, let c_u and c_o respectively denote the cost of underage and overage of one newspaper.

Hence, c_u=p-c

c_o=c

Probability of overage of qth unit of newspaper= P(D<q)=F(q)

Probability of underage of qth unit of newspaper= P(D>q)=1-F(q)

Expected loss on qth unit if overage occurs=c_o.F(q)

= c.F(q)

Expected loss on qth unit if underage occurs

= c_u.{1-F(q)} = (p-c).{1-F(q)}

We assume that the demand distribution is continuous. The cost for order quantity q and demand D is,

Cost(q,D)= c_o(q-D). if D<q

= c_u (D-q), if D>=q

Let f(.) and F(.) denote the density function and distribution of D. Then, expected cost function is given by,

$E[Cost(q,D)] = \int_{D=0}^{\infty} Cost(q,D)f(D)dD$

$= c_o \int_0^q (q-D)f(D)dD +$

$c_u \int_q^{\infty} (D-q)f(D)dD$

In order to find optimal q, we equate the derivative of the expected cost function to 0.

$\frac{dE[Cost(q)]}{dq}$ = c_o F(q) - c_u {1-F(q)}, by Leibnitz rule of differentiation under the sign of integration.

Equating the derivative of cost function to 0, we have,

$$c_o F(q) - c_u \{1-F(q)\} = 0$$

$$\text{or, } F(q) = \frac{cu}{cu+co}$$

$$\text{or, } q = F^{-1}\left(\frac{cu}{cu+co}\right) = q^*, \text{ say. ……..(i)}$$

Also, $\frac{d}{dq}\left(\frac{dE[Cost(q)]}{dq}\right) = c_o F'(q) + c_u F'(q) > 0$ as F is non decreasing, i.e., F'(x) >= 0, for all x.

Testing the second derivative proves that q is a global minimum. Hence, the expected cost is minimised at q^*.

The ratio $\frac{cu}{cu+co}$ is termed as the Critical Fractile, which is the ratio of the cost of underage and the total cost of underage and overage of one newspaper. The Critical Fractile essentially gives the desired service level or the probability of not stocking out. If the demand distribution is known, we can use the critical fractile to find the corresponding order quantity that achieves this level, where the cost to be incurred by the vendor is minimised.

The critical fractile can be derived from another viewpoint. The main motive of the newsvendor is to maximize his profit. Since demand is random, profit is also random.

Note that, the number of units actually sold is given by min(q,D) since if D>q (demand is greater than order quantity or supply), the vendor can sell only what he has in stock, i.e., q. On the other hand, if q>D (supply is greater than demand), the vendor can sell only what is demanded, i.e., D.

So, revenue earned = p.min(q,D)

Cost price of q units of newspaper = cq

Profit = p.min(q,D) – cq

Since D is random, we consider the expectation of profit.

To find the optimal order quantity q that maximizes expected profit, we take the derivative of the expected profit function with respect to q and set it to zero.

$$\frac{dE[profit]}{dq} = \frac{dE[p.\min(q,D)]}{dq} - c \text{ …….(ii)}$$

Now, $E[\min(q,D)] = \int_0^q Df(D)dD + \int_q^\infty qf(D)dD$

Also, $\frac{d}{dq}\left(\int_0^q Df(D)dD\right)$ = q.f(q), by Leibnitz rule of differentiation under the sign of integration,

and $\frac{d}{dq}\left(\int_q^\infty qf(D)dD\right)$

$= 1.\int_q^\infty f(D)dD + q.\frac{d}{dq}\int_q^\infty f(D)dD$

$= \int_q^\infty f(D)dD +$ q.(-f(q)), by Leibnitz rule of differentiation under the sign of integration.

Hence, $\frac{dE[\min(q,D)]}{dq}$ =q.f(q)+ $\int_q^\infty f(D)dD$ - q.f(q)

$= \int_q^\infty f(D)dD$ = 1-F(q)

From (ii),

$\frac{dE[profit]}{dq}$ = p{1-F(q)}-c

Equating $\frac{dE[profit]}{dq}$ to 0, we have,

p{1-F(q)}=c

or, $F(q)=1-\frac{c}{p} = \frac{p-c}{p}$

or, $q=F^{-1}\left(\frac{p-c}{p}\right)$

Also, $\frac{d}{dq}\left(\frac{dE[profit]}{dq}\right) = -pF'(q) < 0$ as F is non decreasing always, i.e., F'(x)>=0, for all x.

So, The critical point determined thus, is the global maxima.

Hence, the point $q = F^{-1}\left(\frac{p-c}{p}\right)$ maximises the expected profit of the vendor and thus is the optimum stocking quantity of newspaper.

Note that $\frac{p-c}{p} = \frac{cu}{cu+co}$.

Comparing this with (i), we note that we get the same optimum stocking quantity in both the methods.

Example: Say, a newsvendor buys newspaper for Rs 1.50 each and sells at Rs 3.50. So, he makes a profit of Rs 2 on selling each newspaper and makes a loss of Rs 1.50 for not being able to sell one newspaper. So, he loses Rs 2 for not being able to sell each newspaper demanded, when he runs out of stock. Thus, we expect the newsvendor to stock some extra newspaper, as turning away a customer incurs him more loss than stocking an extra newspaper. Here, we have,

Selling price of one newspaper, p = Rs 3.50,

Cost price of one newspaper, c = Rs 1.50,

Cost of overage, c_o = Rs(p-c) =Rs 2,

Cost of underage, c_u = Rs c =Rs 1.50

Now, suppose the demand of newspaper follows a Normal Distribution with mean 40 and variance 16. This has been recorded from past demands of newspaper. Our motive is to find the optimum quantity q^* here.

Using the critical fractile formula,

$$q^* = F^{-1}\left(\frac{cu}{cu+co}\right) = F^{-1}\left(\frac{2}{3.50}\right) \approx F^{-1}(0.5714)$$

Then, $F(q^*) = 0.5714$

Or, $P(D < q^*) = 0.5714$

Or, $P\left(\frac{D-40}{4} < \frac{q^*-40}{4}\right) = 0.5714$

Or, $P\left(Z < \frac{q^*-40}{4}\right) = 0.5714$, where Z is a Standard Normal Variate.

Or, $\phi\left(\frac{q^*-40}{4}\right) = 0.5714$

$\frac{q^*-40}{4} \approx 0.18$

Or, $q^* = 40.72 \approx 41$.

Hence, in this case, the optimum stocking quantity is 41 units.

Other Areas of Application of the model:

The newsvendor model can be applied to various other situations, in case the following conditions are satisfied:

- There is a single selling period or a limited time frame for selling the product.
- Demand is uncertain.
- There are costs associated with both overstocking and understocking.

Some of such areas of application are listed below:

a) Food and Flowers: For fresh produce, dairy products, baked goods, and other items with a limited shelf life, a grocery store needs to decide how much milk to stock, knowing that any unsold milk will spoil and become worthless. Florists face uncertain demand for flowers and need to decide how many flowers to order, balancing the risk of running out with the risk of having unsold flowers that will die.

b) Airline Seats: Airlines need to decide how many seats to allocate to each fare class, balancing the risk of empty seats with the risk of turning away potential customers who are willing to pay higher prices.

c) Personal Investments: An investor can use the newsvendor model to decide how much to invest in a risky asset with a random return, balancing the potential for high returns with the risk of losses.

References:

1. Pat DeMarle. "The Basics of the Newsvendor Model". August 27, 2019 https://medium.com/@pdemarle/the-basics-of-the-newsvendor-model-ef756f203433

2. Carol Gameleira. "What is the Newsvendor model?". October 27, 2021 https://supplybrain.ai/en/what-is-the-newsvendor-model/

3. Wendell Burt. "Newsvendor model". https://www.questionai.com/knowledge/kFU0DMoK5W-newsvendor-model

SIMPSON'S PARADOX

Srinjini Bandyopadhyay[1]

B.Sc Statistics Hons, St. Xavier's College (Autonomous), Kolkata

Introduction

Simpson's Paradox is a phenomenon in statistics and probability, which was first described by Edward Hugh Simpson. The name Simpson's paradox was introduced by Colin Ross Blyth. It is also known as the Yule-Simpson effect or reversal paradox or amalgamation paradox. In this phenomenon, a trend which appears in several groups of data, reverses or even disappears when the groups are combined.

What makes Simpson's Paradox paradoxical?
Simpson's Paradox is called a paradox because it often results in conclusions which are paradoxical. There are certain situations where we believe that there is a direct relationship between two variables, but when we consider an additional or third variable, that relationship appears to reverse or disappear.

Illustration of Simpson's Paradox

Let us consider a few examples on combined correlation where Simpson's paradox is observed:

Example 1: Reversal of correlation

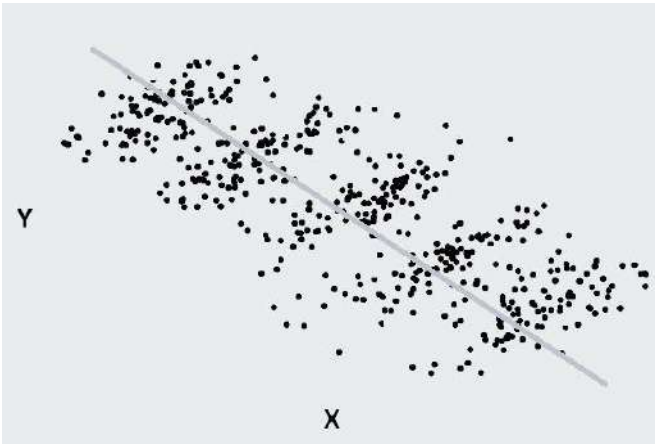


Figure 13.1: Illustration of Simpson's Paradox

In Fig. 13.1, we see that there exists a negative correlation between X and Y as when X increases, Y decreases on an average. However, if we split the data into sub-groups then we see that there is a positive relationship between X and Y as when X increases, Y increases on an average. Thus, we see that the relationship is reversed when the groups are combined.

Example 2: No linear correlation

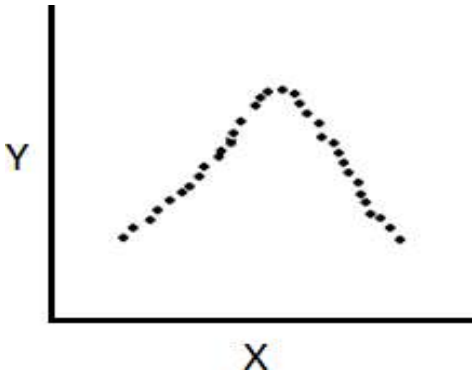


Figure 13.2: Scatterplot showing no linear correlation

In Fig. 13.2, we see that X and Y are not linearly correlated. However, when individual sub- groups are considered then we can see strong positive correlation and strong negative correlation in the first and second half of the graph respectively.

Example 3: Strong positive correlation

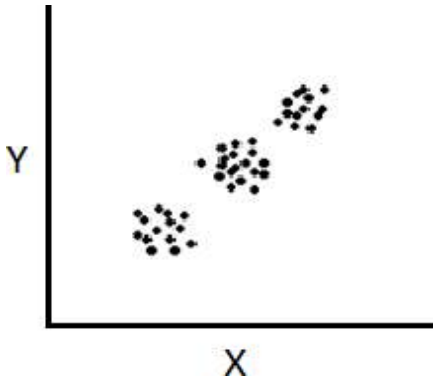


Figure 13.3: Scatterplot showing strong positive correlation

In Fig. 13.3, we see that the combined data shows a strong positive correlation. However, we can observe clustering effect in the individual sub-groups.

3. Importance of the Paradox

This paradox is important because of the following reasons:

a) Statistical relationships are often expected to be immutable. However, they are not always so. The relationship between two variables may increase or decrease or even change direction depending on the variables which are unobserved or being controlled.

b) Through Simpson's paradox, we are often reminded that casual inference, especially in non-experimental studies, can be hazardous. The association observed between two variables might be reversed or eliminated by a third variable which is not observed or controlled.

4. Example of Simpson's Paradox

Suppose there are 4 jars- Jar 1, Jar 2, Jar 3, Jar 4 containing blue and green marbles. Now, let the probability of picking a green marble from Jar 1 be less than that of Jar 2 and the probability of picking a green marble from Jar 3 be less than that of Jar 4. Next, if the contents of Jar 1 and Jar 3 are mixed and the contents of Jar 2 and Jar 4 are mixed, then we expect the probability of picking a green marble from Jar 1+Jar 3 to be less than that of Jar 2+Jar 4. However, Simpson's Paradox says it may not be true.

Now, let us consider 4 boxes, Box 1, Box 2, Box 3 and Box 4. There are 14 blue pens and 2 black pens in Box 1, 4 blue pens and no black pen in Box 2, 2 blue pens and 2 black pens in Box 3, 10 blue pens and 6 black pens in Box 4.

BOX 1	**BOX 2**
Blue – 14 Black – 2	Blue – 4 Black – 0
BOX 3	**BOX 4**
Blue – 2 Black – 2	Blue – 10 Black – 6

Figure 13.4: Content of different boxes before mixing

From the figure, we have,

Probability of picking a blue pen from Box 1=$\frac{14}{16}$

Probability of picking a blue pen from Box 2=$\frac{4}{4}$

Probability of picking a blue pen from Box 3=$\frac{2}{4}$

Probability of picking a blue pen from Box 4=$\frac{10}{16}$

Thus,

Probability of picking a blue pen from Box 1 is less than the probability of picking a blue pen from Box 2 ($\frac{14}{16} < \frac{4}{4}$)

Probability of picking a blue pen from Box 3 is less than the probability of picking a blue pen from Box 4 ($\frac{2}{4} < \frac{10}{16}$)

Next, we mix the pens in Box 1 and Box 3, and the pens in Box 2 and Box 4.

BOX 1 + BOX 3	**BOX 2 + BOX 4**
Blue – 16 Black – 4	Blue – 14 Black – 6

Figure 13.5: Content of different boxes after mixing

After mixing the balls,

Probability of picking a blue pen from (Box 1 + Box 3) = $\frac{16}{20}$

Probability of picking a blue pen from (Box 2 + Box 4) = $\frac{14}{20}$

Thus,
Probability of picking a blue pen from Box 1 + Box 3 is greater than the probability of picking a blue pen from Box 2 + Box 4 ($\frac{16}{20} > \frac{14}{20}$)

Here, we see that initially there is a lower probability of picking blue pens from Box 1 and Box 3 than Box 2 and Box 4 respectively. However, after mixing the content of the boxes, the relationship is reversed as the probability of picking a blue pen from the content of Box 1 and Box 3 is higher than that of Box 2 and Box 4.

5. Spurious Correlation

Simpson's paradox also arises in correlations. Two variables may appear to have a positive correlation even if they have a negative correlation. Such a reversal in correlation is brought about by a lurking confounder.

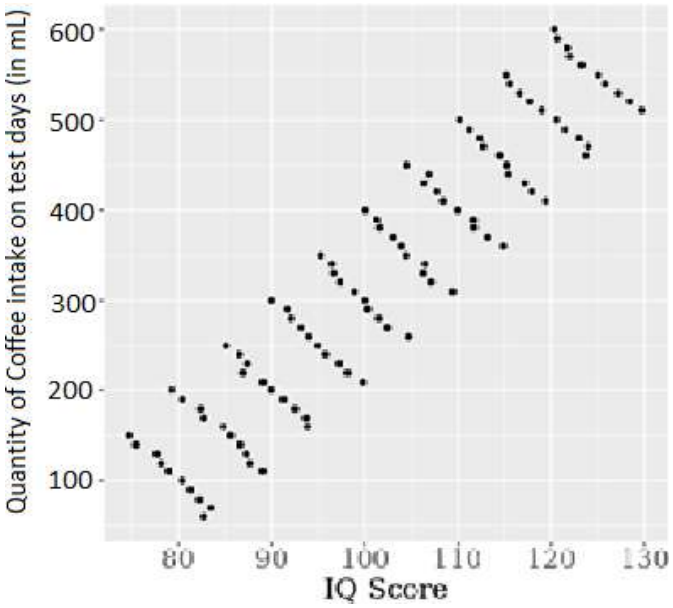


Figure 13.6: Illustration of Simpson's Paradox for bivariate data using a linear regression model

Fig. 13.6 gives the results of coffee intake on performance at an IQ test. We know that coffee intake results in a slight decrease in performance as it makes people less focused. Again, coffee intake co-varies with education level, which in turn co-varies with test performance. When performance is repeatedly measured for different individuals, it is observed that coffee intake has a negative impact on their performance. However, the unconditional regression model of performance as a function of coffee intake misleadingly suggests that coffee consumption strongly improves performance. The reason for the confounding is the causal impact of the hidden covariate, education level, on both coffee consumption and performance.

6. Criticism

Rather than being considered a paradox, Simpson's paradox may actually be considered as a failure to consider casual relationships between two variables or to properly account for the confounding variables. The phenomenon may disappear or even reverse if the data is grouped differently or if the multiple factors affecting the data are considered, suggesting that the Simpson's paradox may cease to be a universal phenomenon.

7. Conclusion

Often, while handling data, there are innumerable factors which operate in the background and influence the response without the experimenter being aware of or interested in them. Such factors can be termed confounding variables. Simpson's paradox can be avoided by taking into consideration these factors while studying the data, and then controlling for these factors in design as well as analysis. Our goal is to determine the true relationship between the independent and outcome variables and obtain unconfounded estimates of treatment effects.

References

[1] https://plato.stanford.edu/entries/paradox-simpson/
[2] https://www.britannica.com/topic/Simpsons-paradox
[3]https://statisticsbyjim.com/basics/simpsons-paradox/

CORE COMMITTEE

Epsilon Delta 2025

Pratiksha Banisitti
Convenor,
εδ '25

Amisha Sengupta
Editor-in-Chief,
Prakarsho Vol XVII

Ankita Sarkar
Co-Convenor,
εδ '25

Srija Upadhyay
Associate
Editor-in-Chief,
Prakarsho Vol XVII

Daniel D. Mondal
Events Head,
εδ '25

Ahan Prodhan
Events Head,
εδ '25

Amissa Parui
Content Head,
εδ '25

Sreyasi Dey
Design Head,
εδ '25

CULTURAL TEAM

From left to right
Aditya Saha, Shrestha Mukherjee, Arshiya Paul, Sumedha Banerjee, Anubhav Hazra.
Raktim Dutta, Soumili Banerjee, Monami Bhattacharyya,
Konkana Adhikary and Koustuv Ghosh.

EVENTS TEAM

From left to right
Soumya Karmakar, Biswanath Das, Soumyadipta Das, Anubhav Hazra,
Daniel D. Mondal, Ahan Prodhan, Arshiya Paul, Sumedha Banerjee,
Akash Roy, Atrijo Roy
Aritra Saha, Rudrajit Chatterjee, Sumeet Sikdar, Ankita Sarkar,
Kankhita Chowdhury, Soumyadeep Saha

MANAGEMENT & FINANCE TEAM

From left to right
Milim C. Lepcha, Zainab Prince, Shrestha Mukherjee,
Konkana Adhikary, Tanwistha Mukherjee, Anushree Ghosh.
Sagnik Roy, Utsab Das, Jayanta Basu, Aritra Saha, Apurba Das

Design and Illustrations by

Sumeet Sikdar and Sreyasi Dey