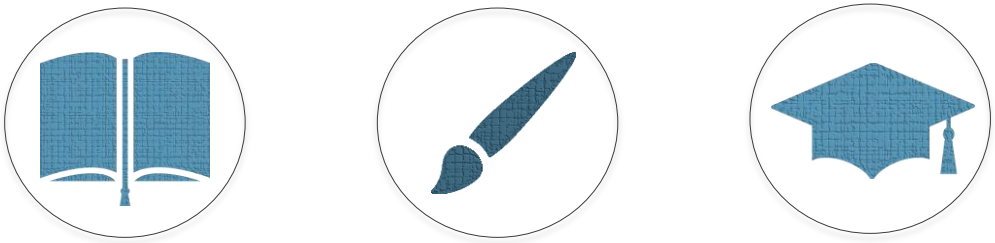


Data Science Concepts

Lesson01A–Probability Theory

Objective

After completing this lesson you will be able to:

- Explain the basic concepts in probability
- Understand the application of empirical probability and Benford's law
- Understand the concepts of probability table and use it for gaining insights

Introduction–Probability

Classical definition of Probability or exact probability: A high school definition with limited application in real life decision making.

- If I have a set of 4 red card and 1 black card, what the probability of a black card coming up if selected from the set of cards randomly shuffled.
- All outcomes are equally likely which is not the case in real life decision making.

Introduction–Probability

Empirical definition of probability: Look at the past data to infer the likelihood of a possible outcome. In a way, we look at the relative frequency of a likely outcome.

- Pharmaceutical companies doing drug test to access the impact of the drug in curing a disease.
- Trying to access the return on investment of a particular investment done in stock exchange.

Introduction–Probability

Subjective definition of probability: In case there is no past history of data, how do you access the likelihood of success or failure. Gut feeling or experience.

- A company trying to access the likelihood of a sale of a new product to be launched into market.

Random Experiment

In probability theory, anything where all the possible outcomes are know but which outcome will appear is uncertain is called a random experiment. Say,

- Rolling a dice, tossing a coin.
- Weather next Monday
- Attrition rate in a company next quarter.
- Stock price of a company next quarter.

Random Experiment

A random experiment will have outcomes

- Basic outcomes: collection of all possible outcomes:
  - Rolling a dice can lead to any of the six numbers on the face.
  - Attrition rate in a company can be any value between 0% to 100%
- Collection of all basic outcomes is called a sample space (S).

We are not interested in all the outcomes but some subset of the sample space. Those are called as events denoted by E(A, B, …X) and may also be termed as random variable

Random variables

A random variable is a numerically valued variable which takes on different values with certain probabilities.

- Random experiment: Attrition rate in an IT company for different years; sample space: S={0%, 1%,…}
  - Observing attrition to be more than 15% may be termed as a random variable. (X >= {15%})

- Random experiment: observe the number of iPhones sold by an Apple store in India in 2018; sample space: S={0,1,2,3,⋯}.
  - Number of iPhones sold between 100, 000 to 250,000 units may be defined as a random variable

Random variables

- Random experiment: Observe cost of treatment at a hospital; sample space: S={1,2,3,⋯}.
  - Cost of treatment exceeding $ 100, 000 may be defined as a random variable

- Random experiment: The number of customers entering a store.
  - Number of customers entering the store on a given day can be defined as a random variable

Probability definitions

- Classical definition of Probability: Assuming all outcomes are equally likely

$$Probability\ of\ event\ A\colon\ P(A) = \frac{number\ of\ basic\ outcome\ that\ satisfies\ A}{Total\ number\ of\ outcome\ in\ sample\ sapce}$$

- Empirical probability definition/relative frequency definition: When proportions are derived from historical data:

$$Probability\ of\ event\ A\colon\ P(A) = \frac{number\ of\ times\ an\ event\ A\ occured\ in\ repeated\ trials}{Total\ number\ of\ trials\ in\ the\ random\ experiment}$$

$$Probability\ of\ event\ A\colon\ P(A) = proportion\ of\ time\ an\ event\ A\ occurs\ in\ large\ number\ of\ trials$$

- Subjective probability definition:

$$Probability\ of\ event\ A\colon\ P(A) = an\ opinion\ or\ belief\ about\ the\ chance\ of\ occurance$$

Fundamental rules of probability– Axioms

- Probability of an outcome in sample space must be 1. $P(S) = 1$
- For any event A, probability of A is between 0 and 1, $0 \leq P(A) \leq 1$
- For disjoint event A and B (event with no elements in common) $\Rightarrow A \cap B = \{\ \} = \emptyset$

$$P(A \cup B) = P(A\ or\ B) = P(A) + P(B)$$

Derived Rules:

- Complement rule: Complement $A^c = All\ the\ elements\ of\ S\ which\ are\ not\ in\ A$

$$P(A^c) = 1 - P(A)$$

- General rule of addition

$$P(A \cup B) = P(A\ or\ B) = P(A) + P(B) - P(A \cap B)$$

Independence of Events

Experiment: Roll two dice

- All possible outcomes: S = {(1,1), (1,2), (1,3),…(4,2), (4,3), (4,4), …(6,4,), (6,5), (6,6)}
- X = Sum of the number which shows up on the face is a random variable. X = {2,3,4,…12}
- What is P(X=4)?

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Sum	Count	Prob
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36
∑	36	1

$$P(X = 4) = P(X = \{(1,3),(2,2),(3,1)\}) = 3/36$$

Independence of Events

If two events are independent then: $P(A \cap B) = P(A \text{ and } B) = P(A) * P(B)$

Roll of two dice is a statistically independent event. Say, sum appearing as 4:

$$\boldsymbol{P(X = 4) = P(X = \{(1,3),(2,2),(3,1)\}) = P(X = \{(1,3) \cup (2,2) \cup (3,1)\})}$$

is the same as

$$\boldsymbol{P(X = 4) = P((1,3)) + P((2,2)) + P((3,1))} \text{ [disjoint]}$$

$$\boldsymbol{P(X = 4) = (P(1) * P(3)) + (P(2) * P(2)) + (P(3) * P(1))} \text{[independence]}$$

$$\boldsymbol{P(X = 4) = \left(\frac{1}{6} * \frac{1}{6}\right) + \left(\frac{1}{6} * \frac{1}{6}\right) + \left(\frac{1}{6} * \frac{1}{6}\right) = \frac{3}{36}}$$

The concept of statistical independence are exploited to simplify the assumptions of Naïve Bayes Classifiers

Application of Probability Theory

Empirical probability and Benford's law

How often do numbers {1,2,3,4,5,6,7,8,9} appear as the first/leading digit in a large dataset?

Empirical probability and Benford's law

How often do numbers {1,2,3,4,5,6,7,8,9} appear as the first/leading digit in a large dataset?

Benford's Law: Chance of observing a lower first digit (1, 2, …) is more than those with a higher first digit (… 8, 9).

Frequency of 'd' being the first digit,

$$P(d) = log_{10}\left(1 + \frac{1}{d}\right), d\epsilon\{1,2,3 \dots 9\}$$

In practice, the accuracy of Benford's Law has been experimented in many fields, such as accounting fraud detection (Nigrini, 1996), electricity bills, stock prices, macroeconomic data like population numbers, death rates, lengths of rivers (Rauch et al. 2011), scientific fraud detection (Diekmann, 2007).

Empirical probability and Benford's law

d	1	2	3	4	5	6	7	8	9
P(d)	.301	.176	.125	.097	.079	.067	.058	.051	.046

When Benford rule generally applies:

- Data should not have pre-defined minimum and maximum value
- It can only be applied to data that fall somewhere between being entirely random (e.g., lottery results) and overly constrained (e.g., the size of new born babies, height of people).

Rule does not apply to: Assigned phone numbers, grades, vehicle registration etc.

Empirical probability and Benford's law

d	1	2	3	4	5	6	7	8	9
P(d)	.301	.176	.125	.097	.079	.067	.058	.051	.046

Useful for fraud detection

Fraudster, to maximize gains, would put in a larger value (starting with 8 or 9) as leading digits and thus would violate benford's law

- Credit card transaction log
- Amount deposited during demonetization in different banks
- Electricity meter reading
- No of votes to different parties in state/general elections (constituency wise)
- Tax returns (A taxpayer compliance application using Benford Law, Nigrni, Mark J (Spring 1996)

Demonstrate Benford law on India Census Data

Conditional Probability

A visit to India, what is the probability that the first person you meet is younger than 21 years?

- Census data shows person residing in India P(0-20 years) = 47.9%
- Census shows that the P(0-20 years | Bangalore) = 40%

Does this information change the above probability?

Conditional Probability

A visit to a city in India, what is the probability that the first person you meet is younger than 21 years?

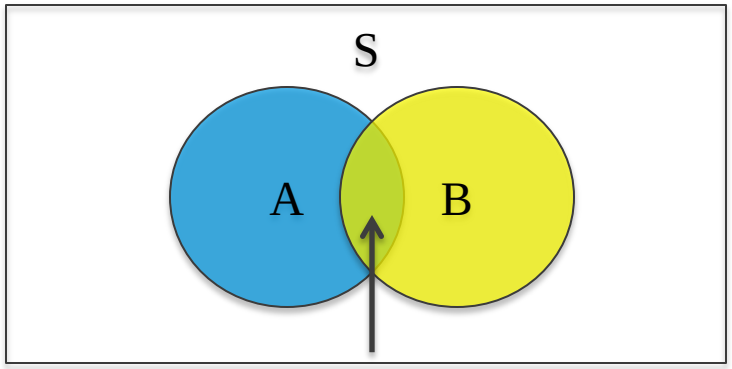
- The person mentions that he is from Bangalore
- Before the occurrence of "Bangalore" P(0-20 years) = 47.9%
- After the occurrence of "Bangalore" P(0-20 years) = 40%

General Idea

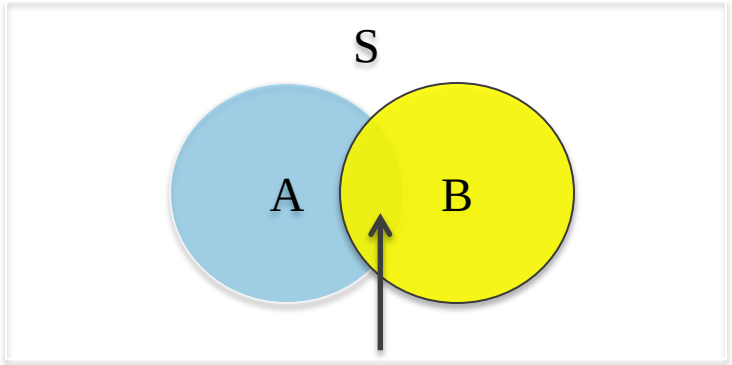
- Probability of event A happening is P(A)
- Information update: B occurred
- What is the probability of event A occurring given event B occurred

$$P(A|B): Conditional\ probability\ of\ A\ given\ B$$

Intuition of conditional probability



$P(A \cap B)$



$P(A \cap B)$

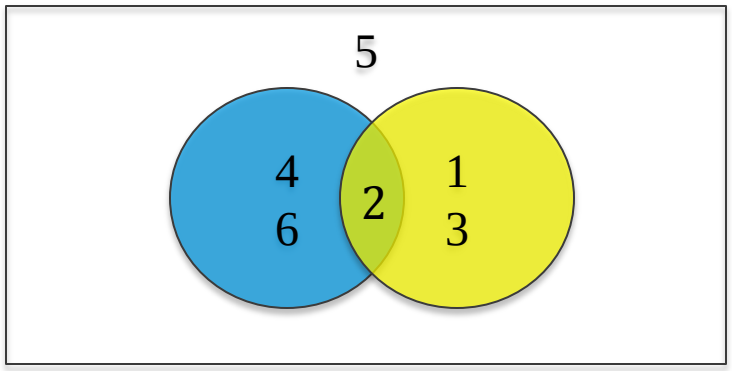
If event B has happened, then the new state space is B. What is left from the probability of event A to happen after B has occurred.

$$P(A|B) = \frac{P(A \cap B)}{\mathrm{P(B)}}$$

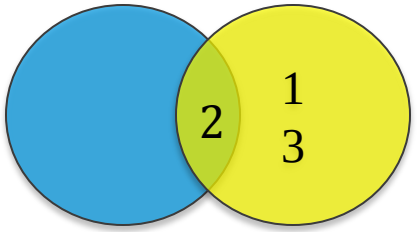
Earlier there was a probability of event A on entire state space S. After B has occurred, we are looking for what is left of event A now just on event B i.e. $P(A \cap B)$

Example

A fair die is rolled



Event B occurred



Event A, getting an even number. Event B, getting the first three numbers.

- A = {2,4,6}; P(A) =1/2
- B= {1,2,3}; P(B) = 1/2
- $(A \cap B) = \{2\}$; $P(A \cap B) = 1/6$ $P(A|B) = 1/3$

- Occurrence of event B changed the probability of event A. Thus A and B are dependent.
- If occurrence of event B does not change the probability of event A. Thus A and B are independent, which means P(A) = $P(A|B)$

General Multiplication Rule

Conditional Probability

- $P(A|B) = \frac{P(A \cap B)}{\mathrm{P(B)}}$
- $P(B|A) = \frac{P(A \cap B)}{\mathrm{P(A)}}$

General Multiplication Rule

- $P(A \cap B) = P(A|B) * P(B)$
- $P(A \cap B) = P(B|A) * P(A)$

In case of Independence of event, $\mathrm{P(A)} = P(A|B)$ or $\mathrm{P(B)} = P(B|A)$

Specialized multiplication rule, in case of independence:

$$P(A \cap B) = P(A) * P(B)$$

Example to apply independence rule

Roll three dice

- Probability of three 1's:

$$P\big((1,1,1)\big) = P(\text{"1"} \cap 1 \cap 1) = P(1) * P(1) * P(1)$$
$$= \frac{1}{6} * \frac{1}{6} * \frac{1}{6} = \frac{1}{216}$$

- Probability of P(X = {5,4,3}):

$$P\big((5,4,3)\big) = P(5 \cap 4 \cap 3) = P(5) * P(4) * P(3)$$
$$= \frac{1}{6} * \frac{1}{6} * \frac{1}{6} = \frac{1}{216}$$

Machinery break down due to heavy load is 1 out of 10 days. With empirical definition of probability: P(Working Machinery) = 0.1. What's the probability of the machine working fine, 2 days in a row? Independence fails!!!

Probability Tables

Foreign visitor to Switzerland

Foreign Visits					
	Africa	Americas	Asia	Europe	totals
hotel	279,870	1,828,085	1,812,706	13,069,622	16,990,283
other	50,481	297,955	203,456	10,790,850	11,342,742
totals	330,351	2,126,040	2,016,162	23,860,472	28,333,025
Calculation of Proportaion					
	Africa	Americas	Asia	Europe	totals
hotel	0.010	0.065	0.064	0.461	0.600
other	0.002	0.011	0.007	0.381	0.400
totals	0.012	0.075	0.071	0.842	1.000

Green colored cells are called marginal probabilities. P(Hotel) = 0.60, P(Africa) = 0.012 etc.

Foreign visitor to Switzerland

Foreign Visits					
	Africa	Americas	Asia	Europe	totals
hotel	279,870	1,828,085	1,812,706	13,069,622	16,990,283
other	50,481	297,955	203,456	10,790,850	11,342,742
totals	330,351	2,126,040	2,016,162	23,860,472	28,333,025
Calculation of Proportaion					
	Africa	Americas	Asia	Europe	totals
hotel	0.010	0.065	0.064	0.461	0.600
other	0.002	0.011	0.007	0.381	0.400
totals	0.012	0.075	0.071	0.842	1.000

Yellow colored cells are called joint probabilities. P(Hotel and Africa) = 0.01, P(Europe and other) = 0.381 etc.

Using the probability of table

What's the probability of a visitor being from Europe given that visitor stays in hotel?

$$P(Europe|Hotel) = \frac{P(Europe \cap Hotel)}{\mathrm{P(Hotel)}} = \frac{0.461}{0.600}$$

What's the probability of a visitor staying in hotel given the visitor is from Europe?

$$P(Hotel|Europe) = \frac{P(Europe \cap Hotel)}{\mathrm{P(Europe)}} = \frac{0.461}{0.842}$$

Probability tables are useful for structuring the information. Things get messier with lot of information in the data.

Set up of probability table

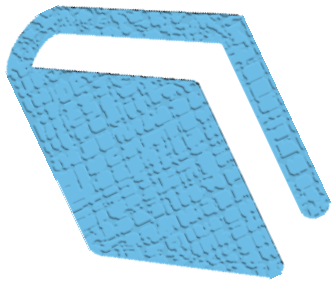
Any number of rows and columns are possible

- Rows are mutually exclusive and completely exhaustive
  - Mutually exclusive means disjoint. Either visitors stayed in hotel or other and there is no third option of stay (completely exhaustive).

- Columns are mutually exclusive and completely exhaustive
  - Mutually exclusive means disjoint. Visitors come from one of these countries and there is no fifth option.(completely exhaustive).

Summary

Summary of the topics covered in this lesson:



- Probability forms the building block for many concepts in statistical learning or inferential learning.
- Probability tables are useful way to gather information and get insights from the data.

End of Lesson01A–Probability Theory