Introduction to Probability

Ms Swapnil Shrivastava
swapnil@cdac.in

Content

- What is Probability?
  - Definitions from different perspectives
- Axioms and Rules of Probability
- Joint Probability
- Marginal Probability
- Independent Events
- Conditional Probability
- Application of probability: Association Rule Mining
- Bayes Theorem
- Summary

Probability

- Most business decisions involve some elements of uncertainty and randomness.
  - Uncertain outcome e.g. computer repair time
  - Assumptions in absence of data e.g. profitability of a new and innovative product
- Probability quantifies uncertainty and is important building block of business analytics application.
- Probability is the likelihood that
  - A new product will be profitable or not
  - A project will complete within 15 weeks
- Probabilities are expressed as values between 0 and 1.

Random Experiment

- An experiment is a process that results in an outcome.
  - Rolling of dice, watching stock market, conducting market research study
- The outcome is not known with certainty.
- Examples:
  - Predicting quarterly revenue of organization
  - Future demand for a product
  - Number of views for an Youtube channel
  - Outcome of a football match
  - Customer churn (likelihood of a customer, no of customers)

Sample Space (S)

- The collection of all possible outcomes of an experiment is called the sample space.
- It is usually represented using the letter S.
- Sample space can be finite or infinite.
- Outcome of a football match
  - S={Win, Draw, Lose}
- Predicting percentage of customer churn
  - $S=\{X \mid X \in \mathit{R}, 0 <= X <= 100)$
- Life of a turbine blade used in an aircraft engine
  - $S =\{X \mid X \in \mathit{R}, 0 <= X <= \infty)$

Event and Outcome

- Event (E) is a subset of a sample space and probability is usually calculated wrt and event.
  - Event A=number of customer churn less than 10
  - Event B=number of customer churn between 10 and 50
  - Event C=number of customer churn exceeding 50
- The outcome of an experiment is a result that we observe.
  - Sum of two dice,
  - change in stock price at the end of week,
  - proportion of consumers who favor the new product

Mathematically

Suppose O_1, O_2, O_3,….O_n are n outcomes where O_i represents the i^{th} outcome in the sample space.

Let $P(O_i)$ be the probability associated with the outcome O_i. Then

$$0 <= P(O_i) <= 1 \text{ for each outcome } O_i$$

$$P(O_1) + P(O_2) + ... + P(O_n) = 1$$

An event is a collection of one or more outcomes from a sample space.

Classical Definition: Theoretical Perspective

- If the process that generates the outcome is known, probabilities can be deduced from theoretical arguments
- Example: Rolling of two dice.
- Possible outcomes: 36
  - (1,1) (1,2) (1,3)(1,4)(1,5)(1,6)
  - (2,1) (2,2) (2,3)(2,4)(2,5)(2,6)
  - (3,1) (3,2) (3,3)(3,4)(3,5)(3,6)
  - (4,1) (4,2) (4,3)(4,4)(4,5)(4,6)
  - (5,1) (5,2) (5,3)(5,4)(5,5)(5,6)
  - (6,1) (6,2) (6,3)(6,4)(6,5)(6,6)
- Probability of rolling a number?
  - Ratio of number of ways of rolling that number to the total number of possible outcomes
- Probability of rolling a 3 is ___

Relative Frequency Definition: Empirical Perspective

- The probability estimation of an event is based on the relative frequency of the occurrence of that event.
- P(X) is the probability of an event X given by

$$P(X) = \frac{\text{Number of observations in favor of event X}}{\text{total number of observations}} = \frac{n(X)}{N}$$

- Say a company has 1000 employees and every year about 200 employees leave the job. Then the probability of attrition of an employee per annum is 200/1000=0.2
- Example: In Computer Repair Times Excel file find the probability that it would be repaired in exactly 10 days.

Subjective Definition: Qualitative perspective

- It is based on judgment and experience.
- Example
  - Financial analysts predicting a 75% chance that the DJIS will increase 10% over the next year.
  - A sports expert might predict, at the start of the football session, a 1-in-5 chance (0.20 probability) of a certain team making it to the Super Bowl.
- In situations where data for analysis is not available.
  - To develop a model to predict the profitability of a new and innovative product.

Algebra of Events

- Assume X, Y and Z are three events of a sample space
- Commutative Rule:
  - $X \cup Y = Y \cup X$ and $X \cap Y = Y \cap X$
- Associative Rule:
  - $(X \cup Y) \cup Z = X \cup (Y \cup Z)$ and $(X \cap Y) \cap Z = X \cap (Y \cap Z)$
- Distributive Rule
  - $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$
  - $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$
- DeMorgan's Law (X^c and Y^c are complementary events of X and Y respectively
  - $(X \cup Y)^c = X^c \cap Y^c$
  - $(X \cap Y)^c = X^c \cup Y^c$

Axioms of Probability

- In 1933, Andrey Kolomogorov, a Russian mathematician laid the foundation.
- The probability of an event E satisfies the following axioms:
  - The probability of event E always lies between 0 and 1. That is $0 <= P(E) <= 1$
  - The probability of the universal set S is 1. That is, $P(S)=1$.
  - $P(XUY)= P(X) + P(Y)$, where X and Y are two mutually exclusive events.

Elementary Rules

- Rule 1: For any event A, the probability of the complement event A^c is $P(A^c) = 1 – P(A)$
- Rule 2: The probability of an empty or impossible event ϕ is zero i.e. $P(\phi)=0$
- Rule 3: If occurrence of an event A implies that an event B also occurs, so that event class A is a subset of event class B, then $P(A)<=P(B)$
- Rule4: The probability that either events A or B occur or both occur is given by $P(A \cup B) = P(A) + P(B) – P(A \cap B)$
- Rule 5: If events A and B are mutually exclusive, then $P(A \cap B)=0$ and hence $P(A \cup B) = P(A) + P(B)$
- Rule 6: If A_1, A_2, ….A_n are n events that form a partition of sample space S, then $P(A_1)+ P(A_2)+…+P(A_n)=1$
- Rule 7: An event is a collection of one or more outcomes from a sample space.

Joint Probability

- Let A and B be two events in a sample space.
- The joint probability of the two events, written as $P(A \cap B)$,is given by
- $P(A \cap B) = \dfrac{\text{number of observations in } A \cap B}{\text{total number of observations}}$
- Example: In a Whatsapp fruit selling group a total of 50 orders were received. Of these 38 ordered mangoes and 20 ordered grapes. Calculate the probability that a customer placed order for both the fruits.

Marginal Probability

- Marginal probability is simply a probability of an event X, denoted by P(X), without any condition.
- In Energy Drink Survey Excel file find the marginal probability for each brand and gender.
- Let
  - G_1: Male
  - G_2 : Female
  - B_1 : Brand1
  - B_2 : Brand2
  - B_3 : Brand3

	Female	Male	Grand Total
Brand 1	9	25	34
Brand 2	6	17	23
Brand 3	22	21	43
Grand Total	37	63	100

- Then Marginal Probabilities are: $P(G_1)$, $P(G_2)$, $P(B_1)$, $P(B_2)$, $P(B_3)$

Independent Events

- Two events A and B are said to be independent when occurrence of one event does not affect the probability of occurrence of the other event.
- It two events A and B are independent then

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

- Example
  - Event A toss won by Indian captain in match 1. Event B toss won by India captain in match 2.
  - Event A is life of an equipment exceeding 100 hours. Event B is life of an equipment exceeding 200 hours.

Conditional Probability

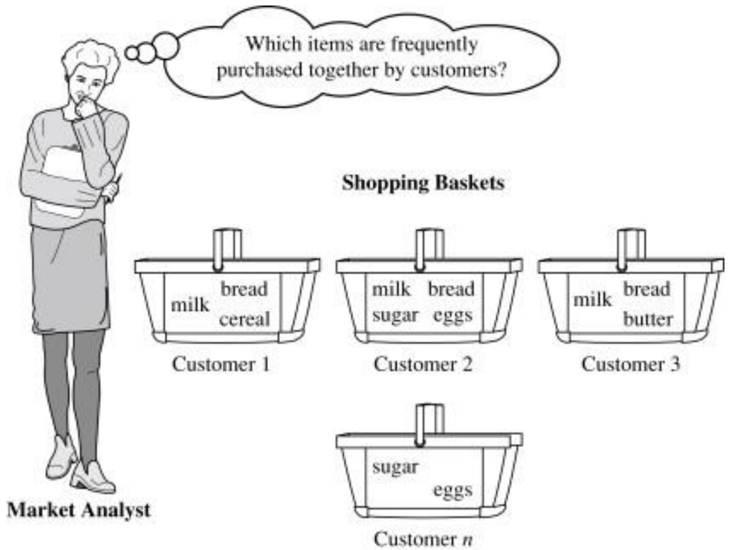
- If A and B are events in a sample, then the conditional probability of an event B given that event A is known to have occurred is denoted by P(B | A).

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0$$

- P(B|A) is read as “the probability of B given A”
  - In Energy Drink Survey Excel file find the probability that an individual is male and prefers brand 1.
- Conditional probabilities are useful in analyzing data in cross-tabulations as well as in other types of applications.
- It can help to predict future purchases based on past purchases.

Association Rule Learning

- It is a popular algorithm used to solve Market Basket Analysis and Recommender System.



- Market Basket Analysis
  - Many customers who buy diapers also buy beer.
  - {Diapers} →{Beer}
- Recommender System

Association Rule Learning

- In a retail context, association rule learning is a method for finding association relationships that exist in frequently purchased items.
- Association Rule is a relationship of the form X→Y i.e. X implies Y.
- X and Y are mutually exclusive sets.
- Association Rules can be created using the Point of Sale (PoS) data from retail stores.
- The strength of association rule is measured in terms of support, confidence and lift.

Point of Sale Data

TID	Items
1	{Apple,Orange,Grapes, Plum,Green Apple, Banana}
2	{Orange, Green Apple, Banana}
3	{Green Apple, Banana}
4	{Apple, Plum}
5	{Apple, Plum,Green Apple, Banana}
6	{Orange, Grapes, Banana}
7	{Orange, Grapes, Banana}

Support

- Support is proportion of times X and Y are purchased together.
- Support s between two sets of products purchased is calculated using the joint probability of those events.

$$\text{Support} = P(X \cap Y) = \frac{n(X \cap Y)}{N}$$

- Where $n(X \cap Y)$ is the number of times both X and Y is purchased together and N is the total number of transactions.

Confidence

- Confidence is the percentage of transactions in the dataset containing Y that also contains X.
- It is the conditional probability of purchasing product Y given the product X is purchased.

$$\text{Confidence} = P(Y|X) = \frac{P(X \cap Y)}{P(X)}$$

- where $P(X \cap Y)$ is the joint probability of X and Y.

Lift

- Lift overcomes one of the disadvantage of using confidence.

$$\text{Lift} = \frac{P(X \cap Y)}{P(X)*P(Y)}$$

- For example, P(X) could be very small, making it less attractive for Market basket Analysis and recommendations among millions of SKUs that a retailer may be selling.

Example

- Association Rules can be generated based on threshold values of support, confidence and lift.
- For example
  - Cut-off for support is 0.25, confidence is 0.5 and lift greater than 1.
  - Assume that X=Apple and Y=Banana. Then
    - Support = 2/7= 0.285
    - Confidence = 2/3 = 0.667
    - Lift = 0.778
  - We can conclude that X implies Y i.e. purchase of apple implies purchase of banana. Since lift is less than 1, rule will be ineffective.

Bayes Theorem

- Consider two events A and B. The two conditional probabilities:

$$P(A|B) = \frac{P(A \cap B)}{P(B\ )} \quad \text{and} \quad P(B|A) = \frac{P(A \cap B)}{P(A\ )}$$

- Using two equations we can show that

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

- It helps to update the probability of an event B when any additional information is provided.
- It helps the decision maker to fine tune his/her belief with every additional data that is received.

Terminology

- P(B) is called the prior probability.
  - Prior probability is estimate of the probability without any additional information
- P(B|A) is called the posterior probability
  - Post the additional information that A has occurred , what is the estimated probability of occurrence of B.
- P(A|B) is called the likelihood of observing evidence A if B is true.
- P(A) is the prior probability of A

Monty Hall Problem

- Let C_1, C_2 and C_3 be the events that the car is behind door 1, 2 and 3 respectively.
- Let D_1, D_2 and D_3 be the events that Monty opens door 1, 2 and 3 respectively.
- Prior probabilities of C_1, C_2 and C_3 are $P(C_1)=P(C_2)=P(C_3)=1/3$

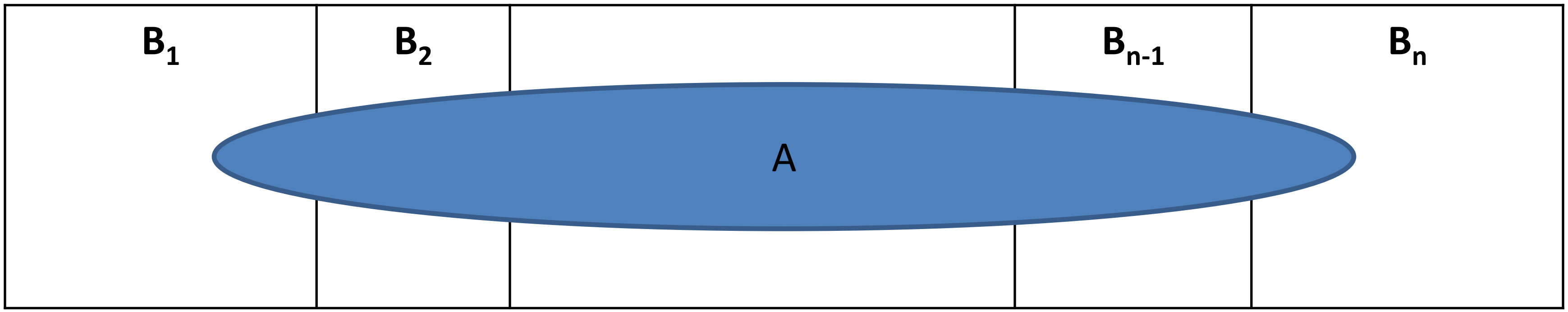
Monty Hall Problem …contd

- Assume that the player has chosen door 1 and Monty opens door 2 to reveal a goat.
- To calculate posterior probability $P(C_1|D_2)$ using Bayes theorem

$$P(C_1|D_2) = \frac{P(D_2|C_1) * P(C_1)}{P(D_2)} = \frac{(1/2)*(1/3)}{(1/2)} = 1/3$$

- $P(C_3|D_2) = 1- P(C_1|D_2) = 1-1/3=2/3$ since $P(C_2|D_2)=0$
- Alternatively using Bayes theorem to solve $P(C_3|D_2)$ also gives 2/3
- If car is behind door 3 and the player has chosen door 1, Monty has to open door 2 with probability 1.

Generalization of Bayes Theorem



$B_1 \cap A$ $B_2 \cap A$ $B_{n-1} \cap A$ $B_n \cap A$

Consider a part manufactured by different suppliers.

Let A denote defective part. P(A) can be written as:

$P(A) = P(A,B_1) + P(A,B_2) + \dots. + P(A,B_n)$

$=P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots.+ P(A|B_n)P(B_n)$

where $P(A,B_1) = P(A \cap B_1)$

Problem

- Black boxes used in aircrafts are manufactured by three companies A, B and C. 75% are manufactured by A, 15% by B, and 10% by C.
- The defect rates of black box manufactured by A, B and C are 4%, 6% and 8%, respectively.
- If a black box tested randomly is found to be defective, what is the probability that it is manufactured by company A?

Summary

- One of the primary objective in analytics is to measure the uncertainty associated with an event or key performance indicator.
- Probability theory is the foundation on which descriptive and predictive analytics models are built.
  - Association Rule Learning
- Probability concepts help in measuring and modelling uncertainty.

References

- James Evans, Business Analytics: Methods, Models and Decisions, Second Edition, Pearson Publication, 2017.
- U Dinesh Kumar, Business Analytics- The Science of Data-Driven Decision Making, Wiley Publication, 2017.