Introduction to Statistical Methods

ISM Team

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Overview of the course & Basic Probability & Statistics (CS -1)

(Session 1: 20th /21st May,2023)

Navneet Goyal
Department of Computer Science & Information Systems
BITS-Pilani, Pilani Campus, India

Overview of the course

- M 1 : Basic Probability & Statistics
- M 2 : Conditional Probability & Bayes' theorem
- M 3 : Probability Distributions
- M 4 : Hypothesis Testing
- M 5 : Prediction & Forecasting
- M 6 : Prediction & Forecasting Gaussian Mixture model & Expectation Maximization

Regression

Time Series analysis

TEXT BOOKS

T1 : Statistics for Data Scientists, An introduction to probability ,statistics and Data Analysis, Maurits Kaptein et al,Springer 2022

T2 : Probability and Statistics for Engineering and Sciences, 8th Edition, Jay L Devore, Cengage Learning

T3 : Introduction to Time Series and Forecasting, Second Edition, Peter J Brockwell, Richard A Davis, Springer.

Choice

Evaluation Components

No	Name	Type	Weight
EC-1(a)	Quizzes – 1 ,2 & 3	Online	10%
	(Best two will be considered)		
EC-1(b)	Assignments - 2	Online	20%
EC-2	Mid-Semester Test	Closed Book	30%
EC-3	Comprehensive Exam	Open Book	40%

Q-1
Ass-1

Q-2
Q-3
Assignment-2

Relative Grading } A | A- | B | B | C | C- | D | E

Repeat this

90 70 60 60 70

EC-1

EC-2

EC-3

Module 1: (Basic Probability & Statistics)

Contact Session	List of Topic Title	Reference
CS - 1	Measures of Central Tendency & Measures of Variability, Data – Symmetric & Asymmetric, outlier detection, 5 point summary, Introduction to probability	T1 & T2

innovate achieve lead

"Statistical thinking will be one day as necessary for efficient citizenship as the ability to read and write"

H G Wells

Statistics

Statistics may be defined as science that is employed to

- Collect the data
- Present and organize the data in a systematic manner
- Analyse the data
- Infer about the data
- Take decision from the data.

In other words, Statistics can also be defined as numerical data with a view to analyse it.

Types of Variable
Qualitative (Categorical): express a qualitative attribute such as hair color, eye color, religion.
Quantitative(Numerical): measured in terms of numbers such as height, weight, number of people.
Nominal: no ordering is possible such as hair color, eye color, religion.
Ordinal: ordering is possible such as health, which can take values such as poor, reasonable, good, or excellent.
Discrete: countable and have a finite number of possibilities such as number of people
Continuous: not countable and have an infinite number of possibilities such as height
INTERVAL: ratio of values of variable do not have any meaning and it does not have an inherently defined zero value such as temperature
RATIO: ratio of values of variable have meaning and it have an inherently defined zero value such as length

Measures of Central Tendency

- Measure of central tendency provides a very convenient way of describing a set of scores with a single number that describes the **PERFORMANCE** of the group.

- Also defined as a single value that is used to describe the "**center**" of the data.

Average

Mea.

- Three commonly used measures of central tendency:
  1. Mean
  2. Median
  3. Mode

Mean

population Mean

sample Mean

$\frac{x_1 + x_2 + x \ldots x_n}{n}$

$\frac{x_1 + x_2 \ldots}{n-1}$

- Also referred as the "**arithmetic average**"
- The most commonly used measure of the center of data
- Numbers that describe what is average or typical of the distribution
- Computation of Sample Mean:

$$\bar{Y} = \frac{\sum Y}{N}$$ = Σ Y/ N = (Y1 + Y2 + Y3 + … Yn) / N where

"Y bar" equals the sum of all the scores, Y, divided by the number of scores, N.

population

sample

- Computation of the Mean for grouped Data

$$\bar{Y} = \frac{\sum f Y}{N}$$ Where f Y = a score multiplied by its frequency

Mean: Grouped Scores

Hours Spent Watching TV	Frequency (f)	fY	Percentage	C%
1	104	104	31.3	31.3
2	130	260	39.2	70.5
3	98	294	29.5	100.0
Total	332	658	100.0	

$$\bar{Y} = \frac{\sum fY}{N} = \frac{658}{332} = 1.98$$

Data of Children watching TV in Bengaluru

Mean

Outliers

6.6

Properties

- It measures stability. Mean is the most stable among other measures of central tendency because every score contributes to the value of the mean.
- It may easily affected by the extreme scores.
- The sum of each score's distance from the mean is zero.
- It can be applied to interval level of measurement
- It may not be an actual score in the distribution
- It is very easy to compute.

Mean depth 2.2 ft

Mean

When to Use the Mean

- Sampling stability is desired.
- Other measures are to be computed such as standard deviation, coefficient of variation and skewness

The Mode

20, 30, 20, 40, 30, 20, 20, 30, 20, 60, 20

mode: 20

mode: 30

- The category or score with the largest frequency (or percentage) in the distribution.

Categorical

Good, Good, Fair, Poor, Good, Fair

- The mode can be calculated for variables with levels of measurement that are: nominal, ordinal, or interval-ratio.

Example:

- Number of Votes for Candidates for Lok Sabha MP. The mode, in this case, gives you the "central" response of the voters: the most popular candidate.

  - Candidate A – 11,769 votes
  - Candidate B – 39,443 votes
  - Candidate C – 78,331 votes

The Mode: "Candidate C"

Mode

Properties

- It can be used when the data are qualitative as well as quantitative.
- It may not be unique.
- It is not affected by extreme values.

When to Use the Mode

- When the “typical” value is desired.
- When the data set is measured on a nominal scale

The Median

- The score that divides the distribution into two equal parts, so that half the cases are above it and half below it.

- The median is the middle score, or average of middle scores in a distribution.
  - Fifty percent (50%) lies below the median value and 50% lies above the median value.
  - It is also known as the middle score or the 50th percentile.

Mean – sensitive
median – X not sensitive

20 25 30 35 400 → Median 30 ✓

20 25 30 35 40 45 → M $\frac{30+35}{2}$: 3

Measures of central tendency

- The mean

$$\frac{x_1 + x_2 + x_3 + \cdots x_n}{n}$$

μ

$n-1$ $\bar{x}$

Drawback?

- the median

10, 15, 20, 25, 26

10, 15, 20, 25, 28, 32

Average of these

- the mode

2, 5, 5, 2, 3, 2, 2, 2

5, 2, 5, 5, 2, 3, 2, 2, 2, 5, 5

Data : Symmetrical and Asymmetrical

Negative Skew

Mean < Median

symmetric

Mean = Median

Bell curve

Mean > Median

Positive Skew

Bell curve

Shape of the distribution of data

- Symmetrical : Mean is equal to median
- Skewed
  - Negatively : mean < median
  - Positively : mean > median
- Bimodal : has two distinct modes
- Multi-modal : has more than 2 distinct modes

symmetric - ✓

not symmetric → normalization → z score ✓

Distribution Shape

Sl. No.	X_1	X_2
1	2	1
2	8	15
3	5	5
4	3	5
5	7	6
6	8	3
7	5	5
8	2	2
9	5	3
Total	45	45

symmetry ✓

Replacing Missing values

Sl. No.	X_1
1	2
2	8
3	5
4	3
5	7
6	8
7	5
8	2
9	5
Total	45

Statistical measures	Group 1
Mean	5
Median	5
Mode	5

Sl. No.	X_2
1	1
2	15
3	5
4	5
5	6
6	3
7	5
8	2
9	3
Total	45

Statistical measures	Group 2
Mean	5
Median	5
Mode	5

Sl. No.	X_1	X_2
1	2	1
2	8	15
3	5	5
4	3	5
5	7	6
6	8	3
7	5	5
8	2	2
9	5	3
Total	45	45

Statistical measures	Group 1 & 2
Mean	5
Median	5
Mode	5

Mean : 2.5

1.2

(2.5 - 1.2, 2.5 - 1.2)

(1.3, 3 7)

Variance / S.D

4 outlier

Measures of central tendency

Mean

Median

Mode

Variance

S.D

Do we need any other measure?

Answer: Yes

$\mu = 20$
$\sigma = 3$

Measures of variability

$(20-3,\ 20+3)$
$(17, 23)$

40

$\frac{(x_i - \mu)^2}{n}$

Three Measures of Variability:

$40 - 20 = 20$

Min = 0
Max = 40

- The Range
- The Variance σ^2
- The Standard Deviations

$\sigma^2 = \frac{\sum (x-\mu)^2}{n} \rightarrow$ variance

$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{n}} \rightarrow$ Standard Deviation

Measure of Variability

Variability can be defined several ways:

- A quantitative distance measure based on the differences between scores
- Describes distance of the spread of scores or distance of a score from the mean

Purposes of Measure of Variability:

- Describe the distribution
- Measure how well an individual score represents the distribution

The Three Measures

Three Measures of Variability:

- The Range
- The Variance
- The Standard Deviations

The Ranges

- The distance covered by the scores in a distribution – From smallest value to highest value
- For continuous data, real limits are used

Range = URL for X_{max} - LRL for X_{min}

- Based on two scores, not all the data – An imprecise, unreliable measure of variability

Example: For a set of scores: 7, 2, 7, 6, 5, 6, 2

Range = Highest Score minus Lowest score = 7 - 2 = 5

The Standard Deviation

- Most common and most important measure of variability is the standard deviation
    - A measure of the standard, or average, distance from the mean
    - Describes whether the scores are clustered closely around the mean or are widely scattered
- Calculation differs for population and samples
- Variance is a necessary *companion concept* to standard deviation but *not the same* concept

The Standard Deviation

Exercise : Find out the deviations of all the data points with the mean….and then find the ‘mean deviation’.

The Standard Deviation

- Mean deviations will always be ‘zero’ !
  (because Mean is a balance point)

Then, how do you find ‘Standard Deviation’ ?

Need a new strategy

The Standard Deviation

New Strategy :

a) First square each deviation score
b) Then sum the Squared Deviations (SS)
c) Average the squared deviations

- Mean Squared Deviation is known as **"Variance"**
- Variability is now measured in squared units

$$Standard\ Deviation = \sqrt{Variance}$$

The Variance

Variance equals mean (average) squared deviation (distance) of the scores from the mean

$$\text{Variance} = \frac{\text{sum of squared deviations}}{\text{number of scores}}$$

where $$SS = \sum (X - \mu)^2$$

The Population Variance

- Population variance equals mean (average) squared deviation (distance) of the scores from the population mean

- Variance is the average of squared deviations, so we identify population variance with a lowercase Greek letter sigma squared: σ 2

- Standard deviation is the square root of the variance, so we identify it with a lowercase Greek letter sigma: σ

Sl. No.	X_1
1	2
2	8
3	5
4	3
5	7
6	8
7	5
8	2
9	5
Total	45

Statistical measures	Group 1
Mean	5
Median	5
Mode	5

Sl. No.	X_1
1	2
2	8
3	5
4	3
5	7
6	8
7	5
8	2
9	5
Total	45

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{45}{5} = 5$$

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}}$$

$$S = \sqrt{\frac{44}{8}} = 2.345$$

μ = population Mean

$= \frac{}{n}$

$\bar{x}$ = Sample Mean

$= \frac{}{n_1 - 1}$

$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}}$

population SD

$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n_1 - 1}}$

Sample S.D

Sl. No.	X_2
1	1
2	15
3	5
4	5
5	6
6	3
7	5
8	2
9	3
Total	45

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{45}{5} = 5$$

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}}$$

$$S = \sqrt{\frac{134}{8}} = 4.093$$

Standard Deviation and Variance for a Sample

- Goal of inferential statistics:
  - Draw general conclusions about population
  -
  - Based on limited information from a sample

- Samples differ from the population
  - Samples have less variability

  - Computing the Variance and Standard Deviation in the same way as for a population would give a biased estimate of the population values

Sample Standard Deviation and Variance

- Sum of Squares (SS) is computed as before
- Formula for Variance has n-1 rather than N in the denominator
- Notation uses s instead of σ

Population of Adult Heights

$$\text{variance of sample} = s^2 = \frac{SS}{n-1}$$

$$\text{standard deviation of sample} = s = \sqrt{\frac{SS}{n-1}}$$

Degrees of Freedom

temp + pressure = 40

$x_1 + x_2 = 40$

$x_2 = 40 - x_1$

ind

dep

- Population variance
  - Mean is known
  - Deviations are computed from a known mean
- Sample variance as estimate of population
  - Population mean is unknown
  - Using sample mean restricts variability

$x_1 + x_2 + x_3 = 1$

$x_3 = 1 - x_1 - x_2$

- Degrees of freedom
  - Number of scores in sample that are independent and free to vary
  - Degrees of freedom (df) = n – 1

Descriptive Statistics

- A standard deviation describes scores in terms of distance from the mean
- Describe an entire distribution with just two numbers (M and s)
- Reference to both allows reconstruction of the measurement scale from just these two numbers
- Means and standard deviations together provide extremely useful descriptive statistics for characterizing distributions

Five point summary of Data

The five number summary of data includes 5 items:

- **Minimum**.
- **Q1** (the first quartile, or the 25% mark).
- **Median**. 50%
- **Q3** (the third quartile, or the 75% mark).
- **Maximum**.

Interquartile range (IQR)

- It is measure of Variation
- Also Known as Midspread : Spread in the Middle 50%
- Difference Between Third & First Quartiles:
- Not Affected by Extreme Values

Interquartile Range = IQR = $Q_3 - Q_1$

Data in Ordered Array: 11 12 13 16 16 17 17 18 21

Position of $Q_1 = \frac{1 \cdot (9 + 1)}{4} = 2.50$, $Q_1 = 12.5$

Position of $Q_3 = \frac{3 \cdot (9 + 1)}{4} = 7.50$, $Q_3 = 17.5$

Interquartile Range = IQR= $Q_3 - Q_1 = 17.5 - 12.5 = 5$

Box-and-Whisker plot

Box-and-Whisker plot

Box-and-Whisker plot

Potential outliers

- The lower limit and upper limit of a data set are given by:

  Lower limit = Q_1 - 1.5 x IQR

  Upper limit = Q_3 + 1.5 x IQR

- Data points that lie below the lower limit or above the upper limit are **potential outliers.**

HW problem :

For the data set below:

82	45	64	80	82	74	79	80	80	78	80	80	48	73	80	79	81	70	78	73

(a.) Obtain and interpret the quartiles.
(b.) Determine and interpret the interquartile range.
(c.) Find and interpret the five-number(point) summary.
(d.) Identify potential outliers, if any.
(e.) Construct and interpret a boxplot.

HW problem :

Human measurements provide a rich area of application for statistical methods. The article "A Longitudinal Study of the Development of Elementary School Children's Private Speech" (*Merrill-Palmer Q.*, 1990: 443–463) reported on a study of children talking to themselves (private speech). It was thought that private speech would be related to IQ, because IQ is supposed to measure mental maturity, and it was known that private speech decreases as students progress through the primary grades. The study included 33 students whose first-grade IQ scores are given here:

82	96	99	102	103	103	106	107	108	108	108	108
109	110	110	111	113	113	113	113	115	115	118	118
119	121	122	122	127	132	136	140	146			

Describe the data and comment on any interesting features.

1) Mean — popul, sampl

2) Median

3) Mode

4) S.D / variance — population, sample

1. Symmetry
2. Missing values
3. Five point summ
4. outliers

$x_1 + temp + prem \le 30$

$x_1 + x_2 \le 30$

$x_2 \le 30 - x_1$

$x_1 + x_2 + x_3 = 20$

$x_3 = 30 - x_1 - x_2$

innovate achieve lead

Introduction to probability

Random Experiment :

❖ The term "**random experiment**" is used to describe any action whose outcome is not known in advance. Here are some examples of experiments dealing with statistical data:

- Tossing a coin
- Counting how many times a certain word or a combination of words appears in the text of the “King Lear” or in a text of Confucius.
- Counting occurrences of a certain combination of amino acids in a protein database.
- Pulling a card from the deck.

the experiment.

Sample spaces and events

- If the experiment consists of a die and we record the outcome, then sample space is S = { 1,2,3,4,5,6}

Event : An event is a subset of sample space of the random experiment.

Definition of probability

Classical approach :

CLASSICAL PROBABILITY

$$\text{Probability of an event} = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

Empirical approach :

Empirical or **relative frequency** is the second type of objective probability. It is based on the number of times an event occurs as a proportion of a known number of trials.

EMPIRICAL PROBABILITY The probability of an event happening is the fraction of the time similar events happened in the past.

In terms of a formula:

$$\text{Empirical probability} = \frac{\text{Number of times the event occurs}}{\text{Total number of observations}}$$

The empirical approach to probability is based on what is called the law of large numbers. The key to establishing probabilities empirically is that more observations will provide a more accurate estimate of the probability.

LAW OF LARGE NUMBERS Over a large number of trials, the empirical probability of an event will approach its true probability.

Axiomatic approach :

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

(2) $0 \leq P(E) \leq 1$

(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \varnothing$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

Thank You

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

innovate achieve lead

BITS Pilani

Session No 2

Axioms of Probability, Probability basics, mutually exclusive and independent events,

(Session 2: 27th/28th MAY 2023)

Contact Session 2

Contact Session 2: Module 1(Module 1:Basic Probability & Statisitcs)

Contact Session	List of Topic Title	Reference
CS - 2	Axioms of Probability,Mutually exclusive and independent events,Problem solving to understand basic probability concepts	T1 & T2
HW	Problems on probability	T1 & T2
Lab		

Agenda

- Experiments, assignment of probabilities
- Events and their probability
- Some basic relationships of probability
- Basic problem solving

RECALL: Random Experiment

Term "**random experiment**" is used to describe any action whose outcome is not known in advance. Here are some examples of experiments dealing with statistical data:

- Tossing a coin
- Counting how many times a certain word or a combination of words appears in the text of the “King Lear” or in a text of Confucius
- counting occurrences of a certain combination of amino acids in a protein database.
- pulling a card from the deck

Sample spaces, sample sets and events

- The ***sample space*** of a random experiment is a set S that includes all possible outcomes of the experiment.

- For example, if the experiment is to throw a die and record the outcome, the sample space is S = { 1,2,3,4,5,6}

- Discrete sample spaces.
- Continuous sample spaces

Event

An event is a subset of the sample space of a random experiment.

An event is a set of outcomes of the experiment. This includes the *null* (empty) set of outcomes and the set of *all* outcomes. Each time the experiment is run, a given event *A* either *occurs*, if the outcome of the experiment is an element of *A*, or *does not occur*, if the outcome of the experiment is not an element of *A*.

Complement of an Event

- The <u>complement</u> of event *A* is defined to be the event consisting of all sample points that are not in *A*.
- The complement of *A* is denoted by A^c.
- The <u>Venn diagram</u> below illustrates the concept of a complement.

Union of Two Events

- The <u>union</u> of events *A* and *B* is the event containing all sample points that are in *A* or *B* or both.
- The union is denoted by *A* U *B*
- The union of *A* and *B* is illustrated below.

Intersection of Two Events

- The <u>intersection</u> of events *A* and *B* is the set of all sample points that are in both *A* and *B*.
- The intersection of *A* and *B* is the area of overlap in the illustration below.

$A \cup B = \{ x \in A \cup B,\ x \in A$ or $x \in B$, $x \in$ both$\}$

$A \cap B = \{ x \in A \cap B,\ x \in A$ and $x \in B \}$

Mutually Exclusive Events

- Two events are said to be mutually exclusive if the events have no sample points in common. That is, two events are mutually exclusive if, when one event occurs, the other cannot occur.

$A \cap B = \phi$

A, B are disjoint sets

A = tossing of coin
B = dice

$A \cap B = \phi$

M.E

A : getting a head
B : getting a tail

Independent events :

Axioms of Probability

Axiom

$0! = 1$

$2! = 2\times1=2$

$3! = 3\times2\times1$

$4! = 4\times3\times2\times1$

$\{m\}$

$\frac{m}{m} = 1 \rightarrow$ certain cent

$\frac{0}{m} = 0$ impossible event

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

(2) $0 \le P(E) \le 1$

(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \varnothing$ → M.E

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

$P(A) = \frac{3}{6}$

R.E: tossing of a coin

$S = \{T, H\}$ → m'

A: getting a head $= n$

$P(A) = \frac{n}{m} = \frac{1}{2}$

R.E: Throwing a die

$S = \{1, 2, 3, 4, 5, 6\}$

A: getting a number ≤ 3

Probability as a Numerical Measure of the Likelihood of Occurrence

Increasing Likelihood of Occurrence

Probability: 0 .5 1

The occurrence of the event is just as likely as it is unlikely.

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

$A \cap B = \phi$ $\quad P(A \cap B) = 0$

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

THE ADDITION RULE

- The probability that event *A* or *B* will occur is given by

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

- If events *A* and *B* are **mutually exclusive**, then the rule can be simplified to

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

Probability and Venn Diagram

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

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

Independent & Dependent

Events are either

- *Independent* (the occurrence of one event has no effect on the probability of occurrence of the other) or
- *Dependent* (the occurrence of one event gives information about the occurrence of the other)

Example

An experiment has the four possible mutually exclusive outcomes A, B, C, D. Check whether the following assignments of probability are permissible:

R.E → A, B, C, D

(a) P(A)= 0.38, P(B) = 0.16, P(C) =0.11, P(D) = 0.35 Permissible

(b) P(A)= 0.31, P(B) = 0.27, P(C) =0.28, P(D) = 0.16 NOT

(c) P(A)= 0.32, P(B) = 0.27, P(C) = -0.06, P(D) = 0.47 NOT

(d) P(A)= 1/2, P(B) = 1/4, P(C) = 1/8, P(D) = 1/16 NOT

(e) P(A)= 5/8, P(B) = 1/6, P(C) = 1/3, P(D) = 2/9 NOT

EXAMPLE: 1

If two dice are thrown , what is the probability that the sum is

a) Greater than 8

b) Less than 6

c) Neither 7 nor 11

{ (1, 1), (1, 2), (1, 3) - - - (1, 6)
(2, 1) (2, 6)
(3, 1) (3, 6)
(4, 1) (4, 6)
(5, 1) - - (5, 6)
(6, 1) - - - (6, 6) }

, P(A) = 36

Solution:

$$P(X = 2) = P\{(1, 1)\} = \frac{1}{36}$$

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

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

$$P(X = 5) = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = \frac{4}{36}$$

$$P(X = 6) = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = \frac{5}{36}$$

$$P(X = 7) = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = \frac{6}{36}$$

$$P(X = 8) = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = \frac{5}{36}$$

$$P(X = 9) = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = \frac{4}{36}$$

$$P(X = 10) = P\{(4, 6)\ (5, 5), (6, 4)\} = \frac{3}{36}$$

$$P(X = 11) = P\{(5, 6), (6, 5)\} = \frac{2}{36}$$

$$P(X = 12) = P\{(6, 6)\} = \frac{1}{36}$$

If two dice are thrown , what is the probability that the sum is

a) Greater than 8

b) Less than 6

c) Neither 7 nor 11

a) $P(\text{sum} > 8)$

$= P(9) + P(10) + P(11) + P(12)$

$= \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36}$

b) $P(\text{sum} < 6)$

$= P(5) + P(4) + P(3) + P(2)$

$= \frac{4}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}$

If two dice are thrown , what is the probability that the sum is

a) Greater than 8

b) Less than 6

c) Neither 7 nor 11

A

$$P(\text{neither } 7 \text{ nor } 11)$$

$$= 1 - [P(7) + P(11)]$$

$$= 1 - \left[\frac{6}{36} + \frac{2}{36}\right]$$

$$= 28/36 = 7/9$$

Example:2

$P(S) = 2/3$

$P(\bar{M}) = 5/9$

$P(M) = 4/9$

The probability that a student passes in statistics examination is 2/3 and the probability that he /she will not pass in mathematics examination is 5/9. The probability that he/she will pass in at least one of the examination is 4/5. Find the probability that he /she will pass in both the examinations

$P(A \cup B)$

$P(S \cup M)$

$P(A \cap B)$

$P(S \cap M)$

$$P(S \cup M) = P(S) + P(M) - P(S \cap M)$$

$$4/5 = 2/3 + 4/9 - P(S \cap M)$$

?

45/144

Solution:

- The probability of passing in statistics P(S)=2/3
- The probability of passing in Mathematics P(M)=1-5/9 =4/9
- Probability of Passing at least one of these examination

$$= P(S \cup M) = 4/5$$

" DATA"

validation

Conc.

- $\boldsymbol{P(S \cup M) = P(S) + P(M) - P(S \cap M)}$.

$$\frac{4}{5} = \frac{2}{3} + \frac{4}{9} - P(S \cap M).$$

$$P(S \cap M) = \frac{2}{3} + \frac{4}{9} - \frac{4}{5} = \frac{14}{45}.$$

The probability of passing both examinations $P(S \cap M) = \frac{14}{45}$

$\frac{45}{14}$

Odd

Example: 3

Suppose that 75% of all investors invest in traditional annuities and 45% of them invest in the stock market. If 85% invest in the stock market and/or traditional annuities, what percentage invest in both?

Solution:

Let A be the event that a randomly selected investor invests in traditional annuities.

Let B be the event that he or she invests in the stock market.

Then P (A) = 0.75, P (B) = 0.45, and P (A ∪ B) = 0.85

Since,

$P(AB) = P(A) + P(B) - P(A \cup B)$

$= 0.75 + 0.45 - 0.85$

$= 0.35.$

$P(A \cap B)$, $P(AB)$ $P(A,B)$ $P(A \cap B)$

Example4

Try

Suppose the manufacturer's specifications for the length of a certain type of computer cable are 2000 ± 10 millimeters. In this industry, it is known that small cable is just as likely to be defective (not meeting specifications) as large cable. That is, the probability of randomly producing a cable with length exceeding 2010 millimeters is equal to the probability of producing a cable with length smaller than 1990 millimeters. The probability that the production procedure meets specifications is known to be 0.99.

(a) What is the probability that a cable selected randomly is too large?

(b) What is the probability that a randomly selected cable is larger than 1990 millimeters?

Solution:

***Solution*:** Let M be the event that a cable meets specifications. Let S and L be the events that the cable is too small and too large, respectively. Then

(a) $P(M) = 0.99$ and $P(S) = P(L) = (1 - 0.99)/2 = 0.005$.

(b) Denoting by X the length of a randomly selected cable, we have

$$P(1990 \leq X \leq 2010) = P(M) = 0.99.$$

Since $P(X \geq 2010) = P(L) = 0.005$,

$$P(X \geq 1990) = P(M) + P(L) = 0.995.$$

This also can be solved by using Theorem 2.9:

$$P(X \geq 1990) + P(X < 1990) = 1.$$

Thus, $P(X \geq 1990) = 1 - P(S) = 1 - 0.005 = 0.995$.

Exercise problems:

Q 1: A Survey conducted by a bank revealed that 40% of the accounts are savings accounts and 35% of the accounts are current accounts and the balance are loan accounts.

$P(C \cap L) = 0 \rightarrow$ Yes

Indept

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

- What is the probability that an account taken at random is a loan account ? 25%
- What is the probability that an account taken at random is **NOT** savings account ? 60%
- What is the probability that an account taken at random is **NOT** a current account 65%
- What is the probability that an account taken at random is a current account or a loan account?

$P(C \cup L) = P(C) + P(L) - P(C \cap L)$

0 ?

$- P(C) \cdot P(L)$

Exercise problems

$P(A \cap \bar{B}) + P(\bar{A} \cap B) = P(A)P(\bar{B}) + P(\bar{A})P(B) = 0.44$

Q 2. A speaks truth in 80% cases and B speaks in 60% cases. What percentage of cases are they likely to contradict each other in stating the same fact.

$\boxed{P(A \cap B) = P(A)P(B)}$

$P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B})$

$P(A \cap \bar{B}) = P(A)P(\bar{B})$

$P(I)$

Q 3. In a certain residential suburb, 60% of all households get Internet service from the local cable company, 80% get television service from that company, and 50% get both services from that company. If a household is randomly selected,

$P(T)$

$P(I \cap T)$

i) What is the probability that it gets at least one of these two services from the company, and

$P(I \cup T) = 60\% + 80\% - 50\% = 90\%$

ii) What is the probability that it gets exactly one of these services from the company?

$\frac{10}{100} + \frac{30}{100} = \frac{40}{100}$ ✓ $P(A \cup B) - P(A \cap B)$

Exercise problems

try

Q 4: The next generation of miniaturised wireless capsules with active locomotion will require two miniature electric motors to manoeuvre each capsule. Suppose 10 motors have been fabricated but that, in spite of test performed on the individual motors 2 will not operate satisfactorily when placed into capsule, to fabricate a new capsule,2 motors will be randomly selected(that is, each pair of motors has the same chance of being selected) find the probability that

a) Both motors will operate satisfactorily in the capsule.

a) One motor will operate satisfactorily and other will not.

Exercise problems

TRY

Q 5. Suppose that 55% of all adults regularly consume coffee, 45% regularly consume carbonated soda, and 70% regularly consume at least one of these two products.

i) What is the probability that a randomly selected adult regularly consumes both coffee and soda?

ii) What is the probability that a randomly selected adult doesn't regularly consume at least one of these two products?

Exercise problems

Q 6. Suppose a student is selected at random from 80 students where 30 are taking mathematics, 20 are taking chemistry and 10 are taking both. Find the probability ‘p’ that the student is taking Mathematics or chemistry?.

Q7: If A and B are events with $P(A \cup B) = 7/8$, $P(A \cap B) = ¼$ and $P(A') = 5/8$, find $P(A)$, $P(B)$ and $P(A \cap B')$**).**

Summary : (Probability)

1. Random Experiment
2. Sample Space
3. Events
4. mutually Exclusive events $P(A \cap B) = 0$
5. Independent event $P(A \cap B) = P(A)P(B)$
6. Probability : $P(A) = \frac{m}{n}$ → favourable cases, n → sample space

7) , $P(S) = 1$

, $0 \le P(A) \le 1$

, A, B are M.E , $P(A \cup B) = P(A) + P(B)$

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

9) ~~P(E~~ Indept A, B $\rightarrow$ $P(A \cap B) = P(A)P(B)$

$P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B})$

$P(A \cap \bar{B}) = P(A)P(\bar{B})$

Conditional Probability :-

S = { }

A : Bread

B : Butter

B : Butter

A : Bread

B.P & diabetic

B | A

B.P → diabetic

diabetic → B.P

A | B . conditional events

diabetic | B.P

B.P | diabetic

conditional events

80%

30%

A → Butter / Bread → (,) ✓

B → Bread / Butter → (,) ✓

B/D

D :

400

diabetic

200 B·P ✓

$P(B/D) = \frac{200}{400}$ P(B)

$P(D) = \frac{400}{1000}$

1000

sample spaces

D/B ✓

200

diabetic

$P(D/B) = \frac{200}{600}$

600

~~diabetic~~

B·P

B

$P(B) = \frac{600}{1000}$

Conditional probabilities

$P(B/D) = \frac{200/1000}{400/1000} = \frac{P(A \cap B)}{P(D)}$

$P(D/B) = \frac{200/1000}{600/1000} = \frac{P(A \cap B)}{P(B)}$

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

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

$P(A/B) = \frac{P(A) \cdot \cancel{P(B)}}{\cancel{P(B)}} = P(A)$

Thanks

events A/B, B/A → same — yes

$A \cap B$, $B \cap$, C — Jam / Butter / Bread

A/B → condition events

$P(A/B)$, $P(B/A)$ → different

$P(\text{Butter}/\text{Bread}) =$

$P(\text{Bread})/\text{Butter} =$

$P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

$P(B/A) = \dfrac{P(A \cap B)}{P(A)}$

$P(A \cap B) = P(A/B)\, P(B)$

$P(A,B) = P(B/A)\, P(A)$

A, B, C $\quad P(A,B,C) = P(C/A,B)\, P(B/A)\, P(A)$

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

innovate achieve lead

BITS Pilani
Pilani Campus

Session 3: Conditional Probability, Independent events and, Total Probability

(Session 3: 3rd/4th June 2023)

RECAPTIULATION:

- Introduction to Probability
- Review of Set Theory
- Counting Principles
- Definition of Probability
- Axioms of Probability
- Addition Rule of Probability

Contact Session 3

Contact Session 3: Module 2(Conditional Probability & Bayes theorem)

Contact Session	List of Topic Title	Reference
CS - 3	Introduction to conditional probability,indepents events, Total probability	T1 & T2
HW	Problems on conditional probability	T1 & T2
Lab		

CONTENTS:

- Conditional Probability

A, B

B/A

- Independent Events

- Total Probability

B/A → (A, B) Jam/Bread → (B, J)

B∩J

A/B → (A, B) Bread/Jam → (B, J)

B∩J

A∩B

A∩B

"conditional events"

CONDITIONAL PROBABILITY

The probabilities assigned to various events depend on what is known about the experimental situation when the assignment is made. Subsequent to the initial assignment, partial information relevant to the outcome of the experiment may become available. Such information may cause us to revise some of our probability assignments

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

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

conditional prob.

$P(\text{Acce} / \text{Mobile}) = 0.75$

$P(\text{Mobile} / \text{Acce}) = 0.20$

CONDITIONAL PROBABILITY

We examine how the information "an event B has occurred" affects the probability assigned to A.

For example, A might refer to an individual having a particular disease in the presence of certain symptoms. If a blood test is performed on the individual and the result is negative , then the probability of having the disease will change (it should decrease, but not usually to zero, since blood tests are not infallible). We will use the notation to represent the conditional probability of A given that the event B has occurred. B is the "conditioning event."

CONDITIONAL PROBABILITY

Let A and B be two events in sample space. The **conditional probability** that event A occurs given that event B has occurred and it is denoted by

$A \cap B \rightarrow A, B$

$B \cap A$

$$P\left(\frac{A}{B}\right) = \frac{P(A \text{ and } B)}{P(B)} \quad \text{OR} \quad P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

It can also written as $P(A \cap B) = P(B)\, P(A/B) \qquad P(B) \neq 0$

$\qquad\qquad\qquad

CONDITIONAL PROBABILITY

Multiplication Rule

In general, $A_1, A_2, \ldots\ldots A_n$ are events in S, then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2/A_1) P(A_3/A_1 \cap A_2) \ldots\ldots P(A_n/A_1 \cap A_2 \cap \ldots\ldots \cap A_{n-1})$$

Examples:1

$P(M)$ $P(M \cap C)$ $P(C)$

In a certain college, 25% of the students failed Maths, 15% of the students failed chemistry and 10% of the students failed both maths and chemistry. A student is selected at random, find

a. If he failed chemistry, what is the probability that he failed Maths?

$$P(M) = 0.25, \quad P(C) = 0.15, \quad P(M \cap C) = 0.1$$

$$P\left(\frac{M}{C}\right) = \frac{P(M \cap C)}{P(C)} = \frac{0.1}{0.15} = 0.6667$$

$P(M|C)$

b. If the failed maths, what is the probability he failed chemistry?

$P(C|M)$

$$P\left(\frac{C}{M}\right) = \frac{P(C \cap M)}{P(M)} = \frac{0.1}{0.25} = 0.4$$

$P(M \cup C) = P(M) + P(C) - P(M \cap C)$

c. What is the probability that the student failed in Maths or chemistry?

$= \frac{25}{100} + \frac{15}{100} - \frac{10}{100} = \frac{30}{100}$

Examples:2

$P(W_P)$ $P(W_P \cap W_B)$

In a housing colony, 70% of the houses are well planned and 60% of the houses are well planned and well built. Find the probability that an arbitrarily chosen house in this colony is well built given that it is well planned.

$P(W_b / W_P) = \dfrac{P(W_b \cap W_P)}{P(W_P)}$

Solution:

Let A be the event that the house is well planned B be the event that the house is well built therefore

$$P(A) = 0.70, \qquad P(A \cap B) = 0.60$$

$$P\left(\frac{B}{A}\right) = \frac{P(B \cap A)}{P(A)} = \frac{0.60}{0.70} = 0.8571$$

Examples: 3

The probabilities of a regularly scheduled flight departs on time is 0.83, arrives on time is 0.82 & it departs and arrives on time is 0.78. Find the probability that a plane

Latest- Past-

(i) arrives on time given that it departed on time,

(ii) departed on time given that it has arrived on time and

Latest- Past-

(iii) find $\boldsymbol{P}\left(\frac{\boldsymbol{A}}{\overline{\boldsymbol{D}}}\right)$ $= \frac{P(A \cap \overline{D})}{P(\overline{D})} = \frac{\quad}{0.17}$

LHS $\frac{P(\bar{A}\cap B)}{P(B)}$

$1 - \frac{P(A\cap B)}{P(B)}$

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

$P(\bar{A}\cap B)$

$P(\bar{A}\cap B)$

Ans:

Let D and A be the events that the flight departs and arrives on time respectively. Then, $\boldsymbol{P(D) = 0.83, P(A) = 0.82}$ and $\boldsymbol{P(D \cap A) = 0.78}$

(i) Probability that the plane arrives on time given that it departed on time is

$$\boldsymbol{P\left(\frac{A}{D}\right) = \frac{P(D\cap A)}{P(D)} = \frac{0.78}{0.83} = 0.9398}$$

(ii) Probability that the plane departed on time given that it has arrived on time is $\boldsymbol{P\left(\frac{D}{A}\right) = \frac{P(D\cap A)}{P(A)} = \frac{0.78}{0.82} = 0.9512}$

(iii) $\boldsymbol{P\left(\frac{A}{\bar{D}}\right) = \frac{P(A\cap \bar{D})}{P(\bar{D})} = \frac{0.82-0.78}{1-0.83} = 0.24}$

This is the probability that the flight arrives on time given that it did not depart on time

$P(\bar{A}) = 1 - P(A)$

$P(\bar{A}/B) = 1 - P(A/B)$

Example: 4

In a housing colony, 70% of the houses are well planned and 60% of the houses are well planned and well built. Find the probability that an arbitrarily chosen house in this colony is well built given that it is well planned.

Solution: Let A be the event that the house is well planned

B be the event that the house is well built

therefore $P(A) = 0.70, \quad P(A \cap B) = 0.60$

$$P\left(\frac{B}{A}\right) = \frac{P(B \cap A)}{P(A)} = \frac{0.60}{0.70} = 0.8571$$

BITS Pilani, Pilani Campus

Example: 5

A news magazine publishes three columns entitled "Art" (A), "Books" (B), and "Cinema" (C). Reading habits of a randomly selected reader with respect to these columns are

Read regularly	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
Probability	.14	.23	.37	.08	.09	.13	.05

Find

A). P(A/B) $\rightarrow \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.06}{0.23}$

B). P(A/B∪C) $\rightarrow \dfrac{P[A \cap (B \cup C)]}{P(B \cup C)} = \dfrac{0.12}{0.47}$

C). P(A/ reads at least one) $P(A/(A \cup B \cup C)) = \dfrac{P[A \cap (A \cup B \cup C)]}{P(A \cup B \cup C)} = \dfrac{0.14}{0.49}$

Solution

We thus have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.08}{.23} = .348$$

$$P(A|B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{.04 + .05 + .03}{.47} = \frac{.12}{.47} = .255$$

$$P(A|\text{reads at least one}) = P(A|A \cup B \cup C) = \frac{P(A \cap (A \cup B \cup C))}{P(A \cup B \cup C)}$$

$$= \frac{P(A)}{P(A \cup B \cup C)} = \frac{.14}{.49} = .286$$

and

$$P(A \cup B|C) = \frac{P((A \cup B) \cap C)}{P(C)} = \frac{.04 + .05 + .08}{.37} = .459$$

INDEPENDENT EVENTS

We can deduce an important result from the conditional probability:

If B has no effect on A, then, $\boldsymbol{P\left(\frac{A}{B}\right) = P(A)}$ Also $\boldsymbol{P\left(\frac{B}{A}\right) = P(B)}$

and we say the events are independent.

i.e., The probability of A does not depend on B.

so, $$\boldsymbol{P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}} = \frac{P(A)\cancel{P(B)}}{\cancel{P(B)}} = P(A)$$

$$\boxed{\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned}}$$

becomes, $$\boldsymbol{P(A) = \frac{P(A \cap B)}{P(B)}}$$

joint prob

or $$\boxed{P(A \cap B) = P(A) \times P(B)}$$

independent events

indep A

indep B

Example: 2

A box contains 20 fuses of which 5 are defective. If two fuses are chosen at random one after the other. What is probability that both the fuses are defective if

(i) the first fuse is replaced,

(ii) (ii) the first fuse is not replaced.

with replacement → Independent

without replacement → Dependent

Solution: Let A be the event that the first fuse is defective and
B be the event that the second fuse is defective

(i) When the first fuse is replaced, the events are independent hence

$$P(A \cap B) = P(A) \times P(B) = \frac{5C_1}{20C_1} \times \frac{5C_1}{20C_1} = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

(ii) When first fuse is not replaced, the events are not independent then

$$P(B \cap A) = P(A) \times P\left(\frac{B}{A}\right) = \frac{5C_1}{20C_1} \times \frac{4C_1}{19C_1} = \frac{1}{19}$$

Miscellaneous examples

1. A box A contains 2 white balls and 4 black balls. Another box B contains 5 white balls and 7 black balls. A ball is transferred from box A to box B, than a ball is drawn from box B. What is the probability that it is white.

2. A person X can hit a target three times in five shots, a second person Y twice in five shots and a third person Z thrice in four shots. They all fire once simultaneously. Find the probability that (i) atleast two shot hit, (ii) exactly two shots hit.

Continued

3. If A and B are two events with P(A)= ½, P(B)= 1/3 and $P(A \cap B) = 1/4$
Find

$$P\left(\frac{A}{B}\right), P\left(\frac{B}{A}\right), P\left(\frac{\overline{A}}{\overline{B}}\right), P\left(\frac{\overline{B}}{\overline{A}}\right), P\left(\frac{A}{\overline{B}}\right)$$

4.Three students X,Y,Z write an examination. Their chances of passing are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ respectively. Find the probability that

(i) all of them pass,

(ii) at least one of them passes,

(iii) at least two of them pass.

Continued

5. The probability that Ram will hit a target is 0.56 and the probability that shyam will hit the target is 0.45. The probability that both will hit the target is 0.25. find the probability that

(i) at least one will hit the target

(ii) exactly one will hit the target.

6. If A and B are two events such that

$$P(A \cap B) = 0.05, P(B) = 0.7, P(A \cap \overline{B}) = 0.02$$

Find (i) $P(A \cup B)$, (ii) $P(A \cup \overline{B})$

Total Probability

Bayes

$P(A_1/S) = 0.2$

$P(A_2/S) = 0.4$

$P(A_3/S) = 0.1$

$P(A_1/S) = \frac{P(S \cap A_1)}{P(S)} = \frac{P(S|A_1)P(A_1)}{\sum_{A_i} P(S|A_i)P(A_i)}$

The Law of Total Probability

Let $A_1, \ldots, A_k$ be mutually exclusive and exhaustive events. Then for any other event B,

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$$

$$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

$A_1 - \frac{70}{100}$ $S/A_1 \to 2\%$

$A_2 - \frac{20}{100}$ $S/A_2 - 3\%$

$A_3 - \frac{10}{100}$ $S/A_3 - 5\%$

$S \to 2\% \text{ of } \frac{70}{100} + 3\% \text{ of } \frac{20}{100} + 5\%$

$= P(S|A_1)P(A_1) + P(S|A_2)P(A_2) + P(S|A_3)P(A_3)$

$S = (S \cap A_1) \cup (S \cap A_2) \cup (S \cap A_3)$

$P(S) = P(S \cap A_1) + P(S \cap A_2) + P(S \cap A_3)$

$= P(S|A_1)P(A_1) + P(S|A_2)P(A_2) + P(S|A_3)P(A_3)$

Metropolitan city

25th floor

Village

Snack.

P(S/A) = 0.8

P(S/A) = 0.2

$S = A_1 \cup A_2 \cup \ldots \cup A_n$ and $A_1 \cap A_2 \cap \ldots \cap A_n = \emptyset$

April Mar | 2020

- Calling Bell
- opened the door

April Mar | 2018

- Calling Bell
- opened the h.

$$\therefore B = B \cap S = B \cap \{A_1 \cup A_2 \cup A_3, \ldots \cup A_n\}$$

Proof:

We have $S = \{A_1 \cup A_2 \cup A_3 \cup \ . \ \ . \ \ . \cup A_n\}$ and $A \subset S$

$$\therefore B = B \cap S = B \cap \{A_1 \cup A_2 \cup A_3, \ . \ \ . \ \ . \cup A_n\}$$

Using distributive law in the R.H.S, we get

Since B ∩ A_i (i = 1 to n) are mutually exclusive, we have by applying addition rule of probability,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots\cdots + P(B \cap A_n)$$

$$i.e., P(B) = \sum_{i=1}^{i=n} P(B \cap A_i)$$

Using multiplication rule on each term on R.H.S, namely

$$P(B \cap A_i) = P(A_i) \cdot P(B \mid A_i) \qquad \text{(A)}$$

$$P(B) = \sum_{i=1}^{i=n} P(A_i) P(B \mid A_i)$$

(B)-Total theorem on Probability

The Theorem of Total Probability (tree diagram)

$P(A/D) = 0.10$

$P(B/D) = 0.50$

$P(C/D) = 0.20$

$\text{Max}\left[P(A/D),\ P(B/D),\ P(C/D)\right]$

$\text{Max}\left[P(D/A)P(A),\ P(D/B)P(B),\ P(D/C)P(C)\right]$

$\Sigma\ X$

MAP Hypothesis

$\text{Max}\left[P(D/A),\ P(D/B),\ P(D/C)\right]$

M.L. Estimator

$\text{Max}\left[\frac{3}{4}, \frac{5}{4}, \frac{6}{4}\right] = 6/4$

Examples :1

D/A —

$P(A) = 35$
$P(B) = 15$
$P(C) = 50$

$P(D/A) = 0.75$
$P(D/B) = 0.95$
$P(D/C) = 0.85$

$P(C/D)$

D / A, B, C — Bayes's theorem

A certain firm has plants A, B, C producing, respectively 35 %, 15% and 50% of the total output. The probabilities of a non – defective product are, respectively, 0.75, 0.95 and 0.85.A Customer receives a ~~bad~~ non-defective product, what is the Chance that product came from the plant C?

$= 0.35 \times 0.75$

$= 0.15 \times 0.95$

$= \dfrac{0.50 \times 0.85}{0.83}$

total prob Σ

Max

$\dfrac{0.35 \times 0.75}{0.83}$ 0.20

$\dfrac{0.15 \times 0.95}{0.83}$ 0.50

$\dfrac{0.50 \times 0.85}{0.83}$ 0.30

Tree Diagram

Solution

Let X: " Customer receives a defective product".

Clearly, P(X)= $P(A)P\left[\frac{X}{A}\right]+P(B)P\left[\frac{X}{B}\right]+P(C)P\left[\frac{X}{C}\right]$

=0.17

Therefore, the chance that product is manufactured by the plant C is

$$P(C\mid X)=\frac{P(C\cap X)}{P(X)} = \frac{0.5\cdot 0.15}{0.17} = 0.4412$$

Example 2

The chances that an academician, a business man and a politician becoming Vice Chancellor of an university are 0.5, 0.3 and 0.2 respectively. The probability that research work will be promoted in the university by these 3 gentlemen are respectively are 0.8, 0.6 and 0.4. It is found Research work has been promoted by the university. What is the chance that an academician has become the VC?

Tree Diagram

Solution

Let **X: " Research work is promoted"**

Clearly, $P(X) = 0.5\times0.8 + 0.3\times0.6 + 0.2\times0.4 = 0.66$

Now to find P[" An Academician is VC" / "Research work is promoted i.e. event X"] $= \frac{0.5\times0.8}{0.66} = 0.6061$

Example 2

The chances that an academician, a business man and a politician becoming Vice Chancellor of an university are 0.5, 0.3 and 0.2 respectively. The probability that research work will be promoted in the university by these 3 gentlemen are respectively are 0.8, 0.6 and 0.4. It is found Research work has been promoted by the university. What is the chance that an academician has become the VC?

Practice problems

Example 1.

An office has 4 secretaries handling respectively 20%, 60%,15% and 5% of the files of all government reports. The probability that they misfile such reports are respectively 0.05, 0.1, 0.1 and 0.05. Find the probability that the misfiled report can be blamed on the first secretary.

Example 2.

In a class 70% are boys and 30% are girls. 5% of boys and 3% of girls are irregular to the classes. What is the probability of a student selected at random is irregular to the classes and what is the probability that the irregular student is a girl?

Continued

Example 3.

Three machines A, B and C produce respectively 60%, 30% and 10% of the total number of items of a factory. The percentage of defective outputs of these machines are 2%, 3% and 4%. An item is selected at random and is found to be defective.

(i) Find the probability that the item was produced by machine C?

(ii) What is the probability that the item was produced by machine C or B?

Example: 4

A card is randomly drawn from an incomplete deck of cards from which the ace of diamonds is missing.

1. What is the probability that the card is "clubs"?
2. What is the probability that the card is a "queen"?
3. Are the events "clubs" and "queen" independent

Example: 5

In a group of children from primary school there are 18 girls and 15 boys. Of the girls, 9 have had measles. Of the boys, 6 have had measles.

1. What is the probability that a randomly chosen child from this group has had measles?

2. If we randomly choose one person from the group of 18 girls, what is the probability that this girl has had measles?

3. Are the events "boy" and "measles" in this example independent?

Webinar - 1 :

Thank you

Quiz - 1 :- 10 questions x 0.5 = 5 Marks

Mid - 30

Compre - 40

Q - 10

A - 20

30

Assignment 1 — Individual

Assignment - 2

Quiz - 2 : After Mi

Quiz - 3

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

innovate achieve lead

BITS Pilani
Pilani Campus

Session 4: Bayes theorem and Naïve Bayes theorem

(Session 4: 10th/11th June 2023)

RECAPTIULATION:

- Introduction to Probability
- Review of Set Theory
- Counting Principles
- Definition of Probability,
- Axioms of Probability,
- Addition Rule of Probability
- Conditional Probability,
- Independent events and,
- Total Probability

Contact Session 4

Contact Session	List of Topic Title	Reference
CS - 4	Bayes theorem(with proof),Introduction to Naïve Bayes concept.	T1 & T2
HW	Problems on Bayes theorem	T1 & T2
Lab	Bayes theorem & Naïve Bayes Concept	Lab 2

Agenda

- Bayes Theorem
- Introduction to Naïve Bayes concept

Text Books

No	Author(s), Title, Edition, Publishing House
T1	Statistics for Data Scientsists, An introduction to probability, statistics and Data Analysis, Maurits Kaptein et al, Springer 2022
T2	Probability and Statistics for Engineering and Sciences, 8th Edition, Jay L Devore, Cengage Learning

BAYES THEOREM:

Let $P = \{A_1,\ A_2,\ A_3,\ \ldots\ , A_k\}$ be a set of exhaustive and mutually exclusive events of a sample space S with $P(A_i) \neq 0$ For each i. If B is any other event associated with A_i with $P(B) \neq 0$, then

$$P\left(\frac{A_i}{B}\right) = \frac{P(A_i)P\left(\frac{B}{A_i}\right)}{P(B)}$$

$$= \frac{P(A_i)P\left(\frac{B}{A_i}\right)}{\sum_{i=1}^{n} P(A_i)P\left(\frac{B}{A_i}\right)}$$

$P(A|B) =$ conditional prob

A_1 or A_2 or A_3 → Bayes theorem

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

$P(E|A_1 \text{ or } A_2 \text{ or } A_3) =$ → Bayes theorem

$P(E|A_1) =$

$P(E|A_2) =$

$P(A_i|E) = \frac{P(A_i \cap E)}{P(E)}$

$= \frac{P(E|A_i)P(A_i)}{\sum P(E|A_i)P(A_i)}$

Geometrical Representation of Bayes theorem

$S = A_1 \cup A_2 \cup \ldots \cup A_n$ and $A_1 \cap A_2 \cap \ldots \cap A_n = \emptyset$

$$\therefore B = B \cap S = B \cap \{A_1 \cup A_2 \cup A_3, \ . \quad . \quad . \cup A_n\}$$

Proof:

The conditional Probability of A_i for any i given B is given as

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

But we know that

$$P(A_i \cap B) = P(A_i) \cdot P(B \mid A_i) \;\&\; P(B) = \sum_{i=1}^{i=n} P(A_i) P(B \mid A_i)$$

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{\sum_{i=1}^{i=n} P(A_i) P(B \mid A_i)}$$

Hence the theorem proved

Example: 1

Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without the disease will show a positive test result only 2% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

Solution:

To use Bayes' theorem, let A_1 = individual has the disease, A_2 = individual does not have the disease, and B = positive test result. Then $P(A_1) = .001$, $P(A_2) = .999$, $P(B|A_1) = .99$, and $P(B|A_2) = .02$. The tree diagram for this problem is in Figure

Next to each branch corresponding to a positive test result, the multiplication rule yields the recorded probabilities. Therefore, $P(B) = .00099 + .01998 = .02097$, from which we have

$$P(A_1 | B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.00099}{.02097} = .047$$

Example: 2

A certain firm has plants A, B, C producing, respectively 35 %, 15% and 50% of the total output. The probabilities of a non – defective product are, respectively, 0.75, 0.95 and 0.85.A Customer receives a bad product, what is the Chance that product came from the plant C?

Tree Diagram

Solution

Let X: " Customer receives a defective product".

Clearly, P(X)= $P(A)P\left[\frac{X}{A}\right]+P(B)P\left[\frac{X}{B}\right]+P(C)P\left[\frac{X}{C}\right]$

=0.17

Therefore, the chance that product is manufactured by the plant C is

$$P(C\,|\,X)=\frac{P(C\cap X)}{P(X)} = \frac{0.5\cdot 0.15}{0.17} = 0.4412$$

Example: 2

The chances that an academician, a business man and a politician becoming Vice Chancellor of an university are 0.5, 0.3 and 0.2 respectively. The probability that research work will be promoted in the university by these 3 gentlemen are respectively are 0.8, 0.6 and 0.4. It is found Research work has been promoted by the university. What is the chance that an academician has become the VC?

A_1 A_2 A_3

R/A_1 R/A_2 R/A_3

$P(A_1/R) =$

$P(A_2/R)$

$P(A_3/R)$

0.5 A_1 — 0.8 R/A_1 $= (0.5)(0.8)$

0.3 A_2 — 0.6 R/A_2 $= (0.3)(0.6)$

0.2 A_3 — 0.4 R/A_3 $= (0.2)(0.4)$

total Σ

$= \dfrac{(0.5)(0.8)}{\Sigma}$

Tree Diagram

Let **X: " Research work is promoted"**

Clearly, $P(X) = 0.5\times0.8 + 0.3\times0.6 + 0.2\times0.4 = 0.66$

Now to find P[" An Academician is VC" / "Research work is promoted i.e. event X"] $= \frac{0.5\times0.8}{0.66} = 0.6061$

BA_1 A_1, A_2, A_3 → ?

Example: A manufacturer of tablets receives its LED screens from three different suppliers, 60% from supplier $B1$, 30% from supplier $B2$, and 10% from supplier $B3$. In other words, the probabilities that any one LED screens received by the plant comes from these three suppliers are 0.60, 0.30, and 0.10. Also suppose that 95% of the LED screens from $B1$, 80% of those from $B2$, and 65% of those from $B3$ perform according to specifications.

1) What is the probability that any one LED screen received by the plant will perform according to specifications?

$P(1/B_1)$

2) Determine the probability that a particular LED screen, which is known to perform according to specifications, came from supplier $B3$.

If A denotes the event that a LED screen received by the plant performs according to specifications, and $B1$, $B2$, and $B3$ are the events that it comes from the respective suppliers

$$A = A \cap [B1 \cup B2 \cup B3]$$
$$= (A \cap B1) \cup (A \cap B2) \cup (A \cap B3)$$
$$= P(A \cap B1) + P(A \cap B2) + P(A \cap B3)$$
$$P(A) = P(B1) \cdot P(A \mid B1) + P(B2) \cdot P(A \mid B2) + P(B3) \cdot P(A \mid B3)$$
$$P(A) = (0.60)(0.95) + (0.30)(0.80) + (0.10)(0.65)$$
$$= 0.875$$

$$P(B3 \mid A) = \frac{(0.10)(0.65)}{(0.60)(0.95) + (0.30)(0.80) + (0.10)(0.65)} = 0.074$$

Example: 3

Example: Two firms V and W consider bidding on a road-building job, which may or may not be awarded depending on the amounts of the bids. Firm V submits a bid and the probability is ¾ that it will get the job provided firm W does not bid. The probability is ¾ that W will bid, and if it does, the probability that V will get the job is only1/3. (a) what is the probability that V will get the job? (b) If V gets the job, what is the probability that W did not bid?

Solution:

Answer: Given $P(V/W^1)=3/4$, $P(W) = 1/3$, $P(V/W)=1/3$ $P(W^1) = 2/3$

(a) $V = (V \cap W) \cup (V \cap W^1) \quad => P(V) = P(V \cap W) + P(V \cap W^1)$

$$P(V) = P(V/W)\,P(W) + P(V/W^1)\,P(W^1) = 11/18$$

(b) $P(W^1/V) = P(V/W^1)\,P(W^1) / P(V) = 9/11$

Suggested problems:

Example . An office has 4 secretaries handling respectively 20%, 60%,15% and 5% of the files of all government reports. The probability that they misfile such reports are respectively 0.05, 0.1, 0.1 and 0.05. Find the probability that the misfiled report can be blamed on the first secretary.

Example . In a class 70% are boys and 30% are girls. 5% of boys and 3% of girls are irregular to the classes. What is the probability of a student selected at random is irregular to the classes and what is the probability that the irregular student is a girl?

Example:

Three machines A, B and C produce respectively 60%, 30% and 10% of the total number of items of a factory. The percentage of defective outputs of these machines are 2%, 3% and 4%. An item is selected at random and is found to be defective. (i) Find the probability that the item was produced by machine C? (ii) What is the probability that the item was produced by machine C or B?

Bayesian Learning

always

- Naive Bayes is a set of simple and efficient machine learning algorithms for solving a variety of classification and regression problems.
- Naive Bayes assumes conditional independence where Bayes theorem does not. This means the relationship between all input features are independent.
- Bayesian learning algorithms that calculate explicit probabilities for hypotheses, such as the naive Bayes classifier, are among the most practical approaches to certain types of learning problems
- For example: Problem of learning to classify text documents such as electronic news articles.
- For such learning tasks, the naive Bayes classifier is among the most effective algorithms known

PV Sindhu Final Gold

10 3

Features of Bayesian learning

- Prior knowledge can be combined with observed data to determine the final probability of a hypothesis.
- Prior knowledge is provided by asserting
  - prior probability for each candidate hypothesis, and
  - probability distribution over observed data for each possible hypothesis.
- New instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities.

Bayes Theorem

- $P(h)$ = prior probability of hypothesis h, before seeing the training data
- $P(D)$ = prior probability of training data D
- $P(h|D)$ = probability of h given D
- $P(D|h)$ = probability of D given h

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Machine Learning

- Generative models
  - Build model to estimate the posterior probability P(Y|X) by estimating
  - likelihood of data given target (hypothesis) P(X|Y)
  - Prior probabilities over target P(Y)
  - In general, for a specific class Y=ck,

$$P(Y = c_k|X) = \frac{P(X|Y = c_k) * P(Y=c_k)}{P(X)}$$

Choosing Hypotheses

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- Generally want the most probable hypothesis given the training data
- *Maximum a posteriori* hypothesis h_{MAP}:

$$h_{MAP} = \arg\max_{h \in H} P(h|D)$$
$$= \arg\max_{h \in H} \frac{P(D|h)P(h)}{P(D)}$$
$$= \arg\max_{h \in H} P(D|h)P(h)$$

P(yes / sunny) = 0.60
P(No / sunny) = 0.40
Max

- If assume $P(h_i) = P(h_j)$ then can further simplify, and choose the *Maximum likelihood* (ML) hypothesis

$$h_{ML} = \arg\max_{h_i \in H} P(D|h_i)$$

Brute Force MAP Hypothesis

- For each hypothesis *h* in *H*, calculate the posterior probability

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

- Output the hypothesis h_{MAP} with the highest posterior probability

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}} P(h|D)$$

MAP Hypothesis

- Using Bayes theorem, we compute the MAP hypothesis for all probable hypothesis (or all unique class labels)
- Identify the best hypothesis describing the data as

$$h_{MAP} = argmax_{h \in H} P(h|D)$$

$$= argmax_{h \in H} \frac{P(D|h)P(h)}{P(D)}$$

$$= argmax_{h \in H} P(D|h)P(h)$$

H: set of all hypothesis

P(D) is independent of h and is same for all hypothesis, therefore dropped

Maximum Likelihood Estimation

- When no prior information is available, all hypothesis are equally likely i.e p(hi) = p(hj)

classifier

prob

non-prob

- This is also true for a balanced class problem where all the classes are equally likely

1. Naive Bayes
2. Logistic regress

1. Decision tree
2. Random For
3. ANN
4. SVM

- This is known as uniform prior
- MAP hypothesis further simplified to
- h_{ML} = argmax P(D/h) (where h belongs to H)

NAÏVE BAYES CLASSIFIER

- Assume independence among attributes X_i when class is given:
  - $P(X_1, X_2, \ldots, X_d \mid Y_j) = P(X_1 \mid Y_j)\, P(X_2 \mid Y_j) \ldots P(X_d \mid Y_j)$
  - Now we can estimate $P(X_i \mid Y_j)$ for all X_i and Y_j combinations from the training data
  - New point is classified to Y_j if $P(Y_j) \prod P(X_i \mid Y_j)$ is maximal.

Depe

Indepe

$y = f(x)$

Input

output

features

response

decision

Regression

numeric

continuous

price → low

1.Fer

Classification

Yes/No

1 -1

or 0/1

0 1

Probabilistic classifier

90% → Diabetes → Yes

20% → Diabetes → NO

pass | fail

Yes/No

Non-Pr

decision Based

Rule Based

Marks ≥ 35 - pass - Yes

< - No - No

Example 1:

classification prob
$y = f(x)$

x = sunny
$y = f(x)$
y — Yes / No

probabilistic classifier

If the weather is sunny, then the player will play or not?

i.e. Play/ Sunny = Yes or No

Note if we know P(Yes/Sunny) and P(No/Sunny) then we can answer the question asked

$$P(Yes/Sunny) = \frac{P(Sunny/Yes)\,P(Yes)}{P(Sunny)}$$

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Steps to Apply Bayes Theorem

Step 1- View or collect "raw" data.

Step 2 - Convert long data to a frequency table

	Play		
	no	yes	Row Total
weather			
Sunny	2	3	5
Overcast	0	4	4
Rainy	3	2	5
Column Total	5	9	14

Step 3 - Row and column sums to get probabilities

Weather probabilities

sunny = 5/14 , rainy = 5/14

Overcast = 4/14

Play probabilities

no = 5/14

yes = 9/14

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Steps to Apply Bayes Theorem

	Play			
	no	yes	Row Total	
weather				
Sunny	2	3	5	P(Sunny)= 5/14
Overcast	0	4	4	P(Overcast) = 4/14
Rainy	3	2	5	P(Rainy)=5/14
Column Total	5	9	14	
	P(no)=5/14	P(yes)=9/14		

Step 4 - Apply probabilities from frequency table to Bayes theorem

$$P(yes \mid sunny) = \frac{P(sunny \mid yes)\, P(yes)}{P(sunny)}$$

$$P(yes|sunny) = \frac{P(yes \cap sunny)}{P(sunny)} = \frac{P(sunny|yes)\,P(yes)}{P(sunny)}$$

weather	No	Yes
Rainy	3	2
sunny	2	3
overcast	0	4
Total	5	9

$5/14 = 0.36$

$5/14 = 0.36$

$4/14 = 0.29$

$5/14 = 0.36$ $\quad$ $9/14 = 0.64$

Now $P(Yes|sunny) = \dfrac{P(Sunny/Yes)\,P(Yes)}{P(sunny)}$

$$= \frac{(3/9)(9/14)}{5/14} = 0.60$$

$$P(No|sunny) = \frac{(2/5)(5/14)}{5/14} = 0.40$$

"Yes"

Conditional independence

- **Definition**: X is conditionally independent of Y given Z, if the probability distribution governing X is independent of the value of Y, given the value of Z

$$(\forall i, j, k)\ P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z_k)$$

$$P(X|Y, Z) = P(X|Z)$$

Example:

$$P(\text{Thunder}|\text{Rain}, \text{Lightning}) = P(\text{Thunder}|\text{Lightning})$$

Applying conditional independence

Naïve Bayes assumes X_i are conditionally independent given Y

e.g., $P(X_1|X_2,Y) = P(X_1|Y)$

$$P(X_1,X_2|Y) = P(X_1|X_2,Y)P(X_2|Y)$$
$$= P(X_1|Y)P(X_2|Y)$$

General form: $P(X_1,\cdots,X_n|Y) = \prod_{j=1}^{n} P(X_j|Y)$

How many parameters to describe $P(X_1,\cdots,X_n|Y)$? $P(\mathrm{Y})$?

Without conditional independence assumption?

With conditional independence assumption?

Naïve Bayes Independence assumption

Assumption:

$$P(X_1, \cdots, X_n | Y) = \prod_{j=1}^{n} P(X_j | Y)$$

i.e., X_i and X_j are <u>conditionally independent</u> given Y for $i \neq j$

Naïve Bayes classifier

- **Bayes rule:**

$$P(Y = y_k | X_1, \cdots, X_n) = \frac{P(Y = y_k) P(X_1, \cdots, X_n | Y = y_k)}{\sum_j P(Y = y_j) P(X_1, \cdots, X_n | Y = y_j)}$$

- **Assume conditional independence among X_i's:**

$$P(Y = y_k | X_1, \cdots, X_n) = \frac{P(Y = y_k) \Pi_i P(X_i | Y = y_k)}{\sum_j P(Y = y_j) \Pi_i P(X_i | Y = y_j)}$$

- **Pick the most probable (MAP) Y**

$$\hat{Y} \leftarrow \underset{y_k}{\operatorname{argmax}} P(Y = y_k) \Pi_i P(X_i | Y = y_k)$$

Prior Probability

MLE

NAÏVE BAYES CLASSIFIER

- Assume independence among attributes X_i when class is given:
  - $P(X_1, X_2, \ldots, X_d \mid Y_j) = P(X_1 \mid Y_j)\, P(X_2 \mid Y_j) \ldots P(X_d \mid Y_j)$
  - Now we can estimate $P(X_i \mid Y_j)$ for all X_i and Y_j combinations from the training data
  - New point is classified to Y_j if $P(Y_j)\, \Pi\, P(X_i \mid Y_j)$ is maximal.

Example 1:

If the weather is sunny, then the player will play or not?

i.e. Play/ Sunny = Yes or No

Note if we know P(Yes/Sunny) and P(No/Sunny) then we can answer the question asked

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Example 2:

$y = f(x)$

$y = f(x_1, x_2, x_3, x_4)$

$x = x_1, x_2, x_3$

If the features of today = (Outlook is Sunny, Temp is Hot, Humidity is Normal, Windy is False), then the player will play or not?

$$P(yes / x) = \frac{P(x / yes)\, P(yes)}{P(x)} = \frac{P(x_1, x_2, x_3, x_4 / yes)\, P(yes)}{P(x_1, x_2, x_3, x_4)}$$

Indep

S. No	Outlook (x_1)	Temp (x_2)	Humidity (x_3)	Windy (x_4)	Play Tennis (y)
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes
10	Sunny	Mild	Normal	False	Yes
11	Rainy	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Sunny	Mild	High	True	No

$P(x_1, x_2, x_3, x_4 / yes) \overset{\text{Indept}}{=} P(x_1/yes) \cdot P(x_2/yes) \cdot P(x_3/yes) \cdot P(x_4/yes)$

$P(x_1, x_2, x_3, x_4) = P(x_1)\,P(x_2)\,P(x_3)\,P(x_4)$

$P(x_1), P(x_2), P(x_3), P(x_4)$

Indept

$P(A, B) = P(A)\,P(B)$

independent

today = (Sunny, Hot, Normal, False)

$X \quad X_1 \quad X_2 \quad X_3 \quad X_4$

yes
no

Bayes + Indept → Naive

solve

Naive Bayes classifier

$$P(yes / X) = \frac{P(X/yes)\,P(yes)}{P(X)}$$

$$\cancel{P(no/X)}$$

$$= \frac{P(x_1, x_2, x_3, x_4 / yes)\,yes}{P(x_1, x_2, x_3, x_4)} = \frac{P(x_1/y)\,P(x_2/y)\,P(x_3/y)\,P(x_4/y)\,P(y)}{P(x_1)\,P(x_2)\,P(x_3)\,P(x_4)}$$

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Example 3:

Magazine Promotion	Watch Promotion	Life Insurance Promotion	Credit Card Insurance	Sex
Yes	No	No	No	Male
Yes	Yes	Yes	Yes	Female
No	No	No	No	Male
Yes	Yes	Yes	Yes	Male
Yes	No	Yes	No	Female
No	No	No	No	Female
Yes	Yes	Yes	Yes	Male
No	No	No	No	Male
Yes	No	No	No	Male
Yes	Yes	Yes	No	Female

$\frac{9}{14}$

NB P( )

Data

New Instance: Magazine Promotion = Yes , Watch Promotion = Yes, Life Insurance Promotion = No, Credit Card Insurance = No then Sex = ?

→ new instance

D = { magazine promotion, watch promotion, Life Insurance promotion, credit card insurance }

h_i = male or Female

$P(\text{male} / \text{Yes, yes, NO, NO}) = ?$

$P(\text{Female} / \text{yes, Yes, NO, NO}) = ?$

	Magazine Promotion		Watch Promotion		L.I Promotion		Credit Card Promotion	
	Male	Female	Male	Female	Male	F	M	F
YES	4	3	2	2	2	3	2	1
NO	2	1	4	2	4	1	4	3
Ratios (YES)	4/6	3/4	2/6	2/4	2/6	3/4	2/6	1/4
NO	2/6	1/4	4/6	2/4	4/6	1/4	4/6	3/4

$P(E_1|male) \cdot P(E_2|male) \cdot P(E_3|male)$

$$P(male|E)$$

$$= \frac{P(E|male)\,P(male)}{P(E)}$$

$$\frac{6}{10} = 3/5$$

$$= \frac{\left(4/6 \cdot 2/6 \cdot 4/6 \cdot 4/6\right)\left(3/5\right)}{P(E)}$$

$$= \frac{0.0593}{P(E)}$$

YES YES NO NO

$$P(\text{Female} / E)$$

$$= \frac{P(E/\text{Female})\, P(\text{Female})}{P(E)}$$

$$= \frac{\left(\frac{3}{4}\right)\left(\frac{2}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) \cdot \frac{2}{5}}{P(E)}$$

$$= \frac{\left(\frac{9}{128}\right)\left(\frac{2}{5}\right)}{P(E)} = \frac{0.0281}{P(E)}$$

$0.0593 > 0.0281$

ie male

Male

Male

Example 4:

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

Condi

Bayes

NB

A: attributes

M: mammals

N: non-mammals

$$P(A \mid M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A \mid N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A \mid M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A \mid N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

Issues with Naïve Bayes Classifier

Naïve Bayes Classifier:

P(Refund = Yes | No) = 3/7
P(Refund = No | No) = 4/7
P(Refund = Yes | Yes) = 0
P(Refund = No | Yes) = 1
P(Marital Status = Single | No) = 2/7
P(Marital Status = Divorced | No) = 1/7
P(Marital Status = Married | No) = 4/7
P(Marital Status = Single | Yes) = 2/3
P(Marital Status = Divorced | Yes) = 1/3
P(Marital Status = Married | Yes) = 0

For Taxable Income:
If class = No: sample mean = 110
sample variance = 2975
If class = Yes: sample mean = 90
sample variance = 25

- P(Yes) = 3/10
  P(No) = 7/10

- P(Yes | Married) = 0 x 3/10 / P(Married)
  P(No | Married) = 4/7 x 7/10 / P(Married)

Issues with Naïve Bayes Classifier

Consider the table with Tid = 7 deleted

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Naïve Bayes Classifier:

P(Refund = Yes | No) = 2/6
P(Refund = No | No) = 4/6
P(Refund = Yes | Yes) = 0
P(Refund = No | Yes) = 1
P(Marital Status = Single | No) = 2/6
P(Marital Status = Divorced | No) = 0
P(Marital Status = Married | No) = 4/6
P(Marital Status = Single | Yes) = 2/3
P(Marital Status = Divorced | Yes) = 1/3
P(Marital Status = Married | Yes) = 0/3
For Taxable Income:
If class = No: sample mean = 91
sample variance = 685
If class = No: sample mean = 90
sample variance = 25

Given X = (Refund = Yes, Divorced, 120K)

P(X | No) = 2/6 X 0 X 0.0083 = 0

P(X | Yes) = 0 X 1/3 X 1.2 X 10^{-9} = 0

Naïve Bayes will not be able to classify X as Yes or No!

Naïve Bayes for Text Classification

- Naïve Bayes is commonly used for text classification
- For a document with k terms $d = (t_1, \ldots, t_k)$

Fraction of documents in c

$$P(c|d) = P(c)P(d|c) = P(c)\prod_{t_i \in d} P(t_i|c)$$

- $P(t_i|c)$= Fraction of terms from all documents in c that are t_i.

Number of times t_i appears in some document in c

$$P(t_i|c) = \frac{N_{ic} + 1}{N_c + T}$$

Laplace Smoothing

Number of unique words (vocabulary size)

Total number of terms in all documents in c

- Easy to implement and works relatively well
- Limitation: Hard to incorporate additional features (beyond words).
  - E.g., number of adjectives used.

A Simple Example:

Text classi → NB → 0 0

Text	Tag
"A great game"	Sports
"The election was over"	Not sports
"Very clean match"	Sports
"A clean but forgettable game"	Sports
"It was a close election"	Not sports

Which tag does the sentence *A very close game* belong to? i.e. P(sports| *A very close game*)

Sports / Not Sports

Feature Engineering: Bag of words i.e use word frequencies without considering order

Using Bayes Theorem:

$$P(\text{sports} \mid \textit{A very close game}) = \frac{P(\textit{A very close game} \mid \text{sports})\, P(\text{sports})}{P(\textit{A very close game})}$$

= P

0

= P(A | sports) P(very | sports) P(close | sports) P(game | sports) · P(...) / P(A) P(very) P(...)

We assume that every word in a sentence is **independent** of the other ones

"close" doesn't appear in sentences of sports tag, So P(close | sports) = 0, which makes product 0 = 0

A Simple Example

Draw | Naive Bayes

Text classification

Text	Tag
"A great game"	Sports
"The election was over"	Not sports
"Very clean match"	Sports
"A clean but forgettable game"	Sports
"It was a close election"	Not sports

Which tag does the sentence "*A very close game*" belong to? i.e. P(sports| *A very close game*)

Feature Engineering: Bag of words i.e use word frequencies without considering order

Using Bayes Theorem:

P(sports| *A very close game*)

$$= \frac{P(\textit{A very close game} \mid \text{sports})\, P(\text{sports})}{P(\textit{A very close game})}$$

We assume that every word in a sentence is **independent** of the other ones

P(A very close game) = P(A) P(very) P(close) P(game)

P(A very closed game | sports) = P(a | sports) P(very | sports) P(close | sports) P(game | sports)

"close" doesn't appear in sentences of sports tag, So P(close | sports) = 0, which makes product 0

$$\frac{P(\text{sports} \mid \text{A very close game})}{P(\text{not sports} \mid \text{A very close game})} = \frac{P(\text{A very close game} \mid \text{sports}) \cdot P(\text{sports})}{P(\text{A very close game})}$$

Laplace smoothing

- Laplace smoothing: we add 1 or in general constant k to every count so it's never zero.
- To balance this, we add the number of possible words to the divisor, so the division will never be greater than 1
- In our case, the possible words are ['a', 'great', 'very', 'over', 'it', 'but', 'game', 'election', 'clean', 'close', 'the', 'was', 'forgettable', 'match'].

Apply Laplace Smoothing

Word	P(word | Sports)	P(word | Not Sports)
a	2+1 / 11+14	1+1 / 9+14
very	1+1 / 11+14	0+1 / 9+14
close	0+1 / 11+14	1+1 / 9+14
game	2+1 / 11+14	0+1 / 9+14

$$P(a|Sports) \times P(very|Sports) \times P(close|Sports) \times P(game|Sports) \times P(Sports)$$
$$= 2.76 \times 10^{-5}$$
$$= 0.0000276$$

$$P(a|Not\,Sports) \times P(very|Not\,Sports) \times P(close|Not\,Sports) \times P(game|Not\,Sports) \times P(Not\,Sports)$$
$$= 0.572 \times 10^{-5}$$
$$= 0.00000572$$

Example :

Doc No	Text
1	I LOVED THE MOVIE
2	I HATED THE MOVIE
3	A GREAT MOVIE ,GOOD MOVIE
4	POOR ACTING
5	GREAT ACTING , A GOOD MOVIE
NEW	I HATED THE POOR ACTING

Example :

Doc No	Text	
1	I LOVED THE MOVIE	_POSITIVE
2	I HATED THE MOVIE	_NEGATIVE
3	A GREAT MOVIE ,GOOD MOVIE	_POSITIVE
4	POOR ACTING	_NEGATIVE
5	GREAT ACTING , A GOOD MOVIE	_POSITIVE
NEW	I HATED THE POOR ACTING	_????

$P(+/x) = ?$

$P(-/x) = ?$

$P(c / x)$

$P(+ / \text{I hated the poor acting}) =$

$P(- / \text{I hated the poor acting}) =$

Based on these probabilities, we can decide the class which the new text belongs

$P(+ \mid I \text{ hated the acting})$

ie $P(c_1 \mid x) = \dfrac{P(x \mid c_1)\, P(c_1)}{\cancel{P(x)}}$

$= P(I \mid +)\, P(\text{hated} \mid +)\, P(\text{the} \mid +)\, P(\text{acting} \mid +)\, P(+)$

$\dfrac{P(I, +)}{P(+)}$ $\qquad$ $\dfrac{P(\text{hated}, +)}{P(+)}$

words	positive	negative
I	1	1
loved	1	0
the	1	1
movie	4	1
hated	0	1
a	2	0
great	2	0
Poor	0	1
acting	1	1
good	2	0

$$P(I \mid +) = \frac{1+1}{14+10}$$

$$P(I \mid -) = \frac{1+1}{6+10}$$

"I hated the poor acting"

word	positive	negative
I	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$
hated	$\frac{0+1}{14+10} = 0.0417$	$\frac{1+1}{6+10} = 0.125$
the	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$
poor	$\frac{0+1}{14+10} = 0.0417$	$\frac{1+1}{6+10} = 0.125$
acting	$\frac{1+1}{14+10} = 0.0833$	$\frac{1+1}{6+10} = 0.125$

x : I hate the Poor acting

$P(+/x)$

$= (\quad)(\quad)(\quad)(\quad)(\quad) \times \underset{\downarrow \atop 3/5}{P(+)}$

$= 6.03 \times 10^{-7}$

$P(-/x)$

$= (\quad)(\quad)(\quad)(\quad)(\quad) \underset{\downarrow \atop 2/5}{P(-1)}$:

$= 1.22 \times 10^{-5}$

$\therefore$ Negative class

Example:

Suppose we got the new message with the words '**Dear Friend**', Decide whether this new message is a normal or spam message?

i.e. Normal/ Dear, Friend = Yes or No

Note if we know P (Normal/ Dear, Friend) and P(Spam/ Dear, Friend) then we can answer the question asked

Email word	Spam	Email word	Spam
Dear	Yes	Friend	No
Friend	No	Friend	Yes
Dear	No	Dear	No
Dear	No	Lunch	No
Dear	No	Friend	No
Friend	No	Dear	No
Lunch	No	Dear	No
Friend	No	Dear	No
Lunch	No	Dear	No
Dear	Yes	Money	Yes
Money	Yes	Money	No
Money	Yes	Money	Yes

Steps to Apply Bayes Theorem

Step 1- View or collect "raw" data.

Step 2 - Convert long data to a frequency table

	Play		
	normal	spam	Row Total
word			
Dear	8	2	10
Friend	5	1	6
Lunch	3	0	3
Money	1	4	5
Column Total	17	7	24

Step 3 - Row and column sums to get probabilities

Email word	Spam	Email word	Spam
Dear	Yes	Friend	No
Friend	No	Friend	Yes
Dear	No	Dear	No
Dear	No	Lunch	No
Dear	No	Friend	No
Friend	No	Dear	No
Lunch	No	Dear	No
Friend	No	Dear	No
Lunch	No	Dear	No
Dear	Yes	Money	Yes
Money	Yes	Money	No
Money	Yes	Money	Yes

Steps to Apply Bayes Theorem

Play word	normal	spam	Row Total
Dear	8	2	10
Friend	5	1	6
Lunch	3	0	3
Money	1	4	5
Column Total	17	7	24

As P (N/ D, F) > P (S/ D, F) , we can decide that Dear Friend is Normal message.

Step 4 - Apply probabilities from frequency table to Bayes theorem

$$P(N/D,F) = \frac{P(D,F/N).P(N)}{P(D,F)} = \frac{P(D/N).P(F/N).P(N)}{P(D).P(F)} = \frac{\left(\frac{8}{17}\right).\left(\frac{5}{17}\right).\left(\frac{17}{24}\right)}{\left(\frac{10}{24}\right).\left(\frac{6}{24}\right)} = \frac{0.098}{0.104} = 0.9423$$

$$P(S/D,F) = \frac{P(D,F/S).P(S)}{P(D,F)} = \frac{P(D/S).P(F/S).P(S)}{P(D).P(F)} = \frac{\left(\frac{2}{7}\right).\left(\frac{1}{7}\right).\left(\frac{7}{24}\right)}{\left(\frac{10}{24}\right).\left(\frac{6}{24}\right)} = \frac{0.012}{0.104} = 0.1153$$

Example continued:

Suppose we got the new message contains the word **'Lunch Money Money Money Money'** , Decide whether this new message is a normal or spam message?

Email word	Spam	Email word	Spam
Dear	Yes	Friend	No
Friend	No	Friend	Yes
Dear	No	Dear	No
Dear	No	Lunch	No
Dear	No	Friend	No
Friend	No	Dear	No
Lunch	No	Dear	No
Friend	No	Dear	No
Lunch	No	Dear	No
Dear	Yes	Money	Yes
Money	Yes	Money	No
Money	Yes	Money	Yes

Steps to Apply Bayes Theorem

Play	normal	spam	Row Total
word			
Dear	8	2	10
Friend	5	1	6
Lunch	3	0	3
Money	1	4	5
Column Total	17	7	24

We can observe that we have to classify any message with Lunch as Normal message, no matter how many times we see the word Money and that's the problem.

To work around this problem add 1 count to the frequency table to each word(Laplace smoothing)

Step 4 - Apply probabilities from frequency table

$$P(N).\ P(L/N).\ P(M/N)^4 = \left(\frac{17}{24}\right).\left(\frac{3}{17}\right).\left(\frac{1}{17}\right)^4 = 0.0000015$$

$$P(S).\ P(L/S).\ P(M/S)^4 = \left(\frac{7}{24}\right).\left(\frac{0}{7}\right).\left(\frac{4}{7}\right)^4 = 0$$

Steps to Apply Bayes Theorem

Play word	normal	spam	Row Total
Dear	8+1	2+1	12
Friend	5+1	1+1	8
Lunch	3+1	0+1	5
Money	1+1	4+1	7
Column Total	21	11	32

Step 4 - Apply probabilities from frequency table

$$P(N).\ P(L/N).\ P(M/N)^4 = \left(\frac{21}{32}\right).\left(\frac{4}{21}\right).\left(\frac{2}{21}\right)^4 = 0.00001$$

$$P(S).\ P(L/S).\ P(M/S)^4 = \left(\frac{11}{32}\right).\left(\frac{1}{11}\right).\left(\frac{5}{11}\right)^4 = 0.00133$$

As $P(S/L, M^4) > P(N/L, M^4)$, we can decide that **Lunch Money Money Money Money is Spam** message.

Naïve Bayes Classifier Applications

MLP → NB

NB

Deep Learning

Categorizing News

Email Spam Detection

Face Recognition

Sentiment Analysis

NB

Naive Bayes Classifier

- ✓ Along with decision trees, neural networks, one of the most practical learning methods.
- ✓ When to use
  - ✓ Moderate or large training set available
  - ✓ Attributes that describe instances are conditionally independent given classification
- ✓ Successful applications:
  - ✓ Diagnosis
  - ✓ Classifying text documents

Learning to Classify Text

- Why?
  - Learn which news articles are of interest
  - Learn to classify web pages by topic

- Naive Bayes is among most effective algorithms
- What attributes shall we use to represent text documents??

Baseline: Bag of Words Approach

aardvark	0
about	2
all	2
Africa	1
apple	0
anxious	0
...	
gas	
...	
oil	
…	
Zaire	

Practical Issues of Bayesian learning

- Require initial knowledge of many probabilities
  - Often estimated based on background knowledge, previously available data, and assumptions about the form of the underlying distributions.
- Significant computational cost required to determine the Bayes optimal hypothesis in the general case (linear in the number of candidate hypotheses)

HW: Exercise

Consider the car theft problem with attributes Color, Type, Origin, and the target, Stolen can be either Yes or No.

we need to classify whether the car is stolen, given the features of the car.

Given the Red color Domestic SUV car Find the probability of whether the car is stolen?

Color	Type	Origin	Stolen?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

HW: Exercise

BNB

If the weather is Snowy, then the player will play or not?

NB → GNB

Gaussian Naive Bayes

GNB

weather	Player play
Sunny	yes
Rainy	no
Cloudy	yes
Sunny	no
Sunny	yes
snowy	no
Rainy	yes
Cloudy	no
Cloudy	yes
Sunny	yes
snowy	no
Cloudy	yes
Rainy	no
snowy	no
snowy	yes

A/B , B/A

3 coins → {HHH, THT, HHT, ... TTT}

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

P(1 head) = 3/8

P(2 head) = 3/8

P(3 head) =

random

Thanks

X = number of head

X = 0

X = 1

X = 2

X = 3

→ Random variable

Y = no of tails

Random variable

X = 0, 1, 2, 3 — Discrete

X = [-10, 10] — Continuous