e-PGPathshala
Subject : Computer Science
Paper: Data Analytics
Module No 13: CS/DA/13- Data Analysis Foundations – Probabilistic and Bayesian Approaches
Quadrant 1 – e-text

1.1 Introduction

The word **probability** is used to mean the chance that a particular **event** (or set of events) will occur expressed on a linear scale from 0 to 1.This chapter covers the fundamentals of probability and probabilistic analytics with real life examples and an overview on Bayesian theory and its application in analytics.

1.2 Learning Outcomes

- Learn the fundamentals of probability
- Understand the probabilistic analytics with examples
- Learn the application of Bayes' logic for analysis

1.3 Discrete Random variable

A **discrete variable** is a variable which can only take a countable number of values. For a discrete random variable, its **probability distribution** (also called the **probability distribution function**) is any table, graph, or formula that gives each possible value and the probability of that value.

For example, if a coin is tossed three times, the number of heads obtained can be 0, 1, 2 or 3. The probabilities of each of these possibilities can be tabulated as shown:

Number of Heads	0	1	2	3
Probability	1/8	3/8	3/8	1/8

In this example, the number of heads can only take 4 values (0, 1, 2, 3) and so the variable is discrete. The variable is said to be **random** if the sum of the probabilities is one. Other examples may include, a variable A = The PM of India 2020 will be female, A = You wake up tomorrow with a headache, A = You have swineflue, etc.

1.4 Sample space

A sample space is a set of all possible outcomes for an activity or experiment and the elements of the sample space are called outcomes. Few more examples are given in table given below:

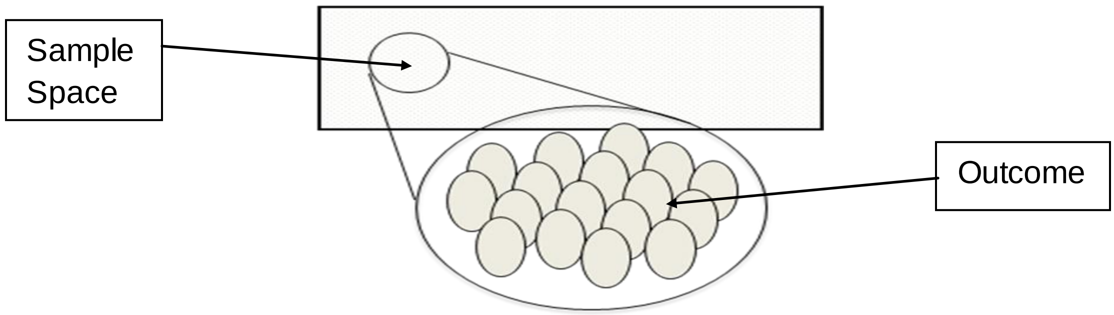
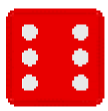
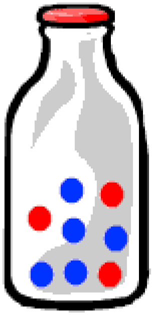


Figure 1. Sample and Sample Space

Assume you are a doctor. This is the sample space of “patients you might see on any given day”. The outcomes may be like Non-smoker, female, diabetic, headache, Smoker, male, herniated disk, back pain, mildly schizophrenic, delinquent medical bills, etc.

Table1. Examples for activities and sample space

Activity	Sample Space
Rolling a die 	There will be 6 outcomes in the sample space: **{1, 2, 3, 4, 5, 6}**
Tossing a coin 	There will be 2 outcomes in the sample space: **{ Heads, Tails}**

Drawing a card from a standard deck 	There will be 52 cards in the sample space: {Spades: 2,3,4,5,6,7,8,9,10, ace, jack queen, king, Clubs: 2,3,4,5,6,7,8,9,10, ace, jack, queen, king, Diamonds: 2,3,4,5,6,7,8,9,10, ace, jack, queen, king, Hearts: 2,3,4,5,6,7,8,9,10, ace, jack, queen, king}
Drawing one marble from the bottle 	There will be 8 marbles in the sample space: **{blue marble, blue marble, blue marble, blue marble, blue marble, red marble, red marble, red marble}**
Rolling a pair of dice 	There will be 36 outcomes in the sample space: **{(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)}**
Choosing an outfit from a green blouse, a red blouse, a black skirt, a pair of sneakers, and a pair of sandals	There will be 4 outfit combinations in the sample space: **{green blouse-skirt-sneakers, green blouse-skirt-sandals, red blouse-skirt-sneakers, red blouse-skirt-sandals}**

A sample spaces can become very large. When determining a sample space, be sure to consider ALL possibilities. Oftentimes this can be a difficult task. To make this process easier, we may wish to use the **Counting Principle**.

The Fundamental Counting Principle: If there are **a** ways for one activity to occur, and **b** ways for a second activity to occur, then there are **a • b** ways for both to occur.

Examples:

1. Activities: roll a die and flip a coin

   There are 6 ways to roll a die and 2 ways to flip a coin.

   There are 6 • 2 = 12 ways to roll a die and flip a coin.

2. Activities: draw two cards from a standard deck of 52 cards without replacing the cards

   There are 52 ways to draw the first card. There are 51 ways to draw the second card. There are 52 • 51 = 2,652 ways to draw the two cards.

3. The Counting Principle also works for more than two activities: A coin is tossed five times and there are 2 ways to flip each coin

   There are 2 • 2 • 2 • 2 • 2 = 32 arrangements of heads and tails

4. A die is rolled four times and here are 6 ways to roll each die. There are 6 • 6 • 6 • 6 = 1,296 possible outcomes.

1.5 Events and Outcomes

a) Events

In **probability** theory, an **event** is a set of outcomes of an experiment (a subset of the sample space) to which a **probability** is assigned. When we say "Event" we mean one (or more) outcomes.

Example Events:

1. Getting a Tail when tossing a coin is an event
2. Rolling a "5" is an event.
3. An event can include several outcomes:
4. Choosing a "King" from a deck of cards (any of the 4 Kings) is also an event
5. Rolling an "even number" (2, 4 or 6) is an event

Events can be of the following type:

a. **Independen**t (each event is not affected by other events),

b. **Dependen**t (also called "Conditional", where an event is affected by other events)

c. **Mutually Exclusive** (events can't happen at the same time)

Independent Events:

Events can be "Independent", meaning each event is not affected by any other events.

This is an important idea! A coin does not "know" that it came up heads before ... each toss of a coin is a perfect isolated thing.

Example: You toss a coin three times and it comes up "Heads" each time ... what is the chance that the next toss will also be a "Head"?

The chance is simply 1/2, or 50%, just like ANY OTHER toss of the coin.

What it did in the past will not affect the current toss!

Some people think "it is overdue for a Tail", but really truly the next toss of the coin is totally independent of any previous tosses.

Saying "a Tail is due", or "just one more go, my luck is due" is called The Gambler's Fallacy.

Dependent Events

But some events can be "dependent" ... which means they can be affected by previous events.

Example: Drawing 2 Cards from a Deck

After taking one card from the deck there are less cards available, so the probabilities change!

Let's look at the chances of getting a King. For the 1st card the chance of drawing a King is 4 out of 52.

But for the 2nd card:

If the 1st card was a King, then the 2nd card is less likely to be a King, as only 3 of the 51 cards left are Kings.

If the 1st card was not a King, then the 2nd card is slightly more likely to be a King, as 4 of the 51 cards left are King.

This is because we are removing cards from the deck.

Replacement: When we put each card back after drawing it the chances don't change, as the events are independent.

Without Replacement: The chances will change, and the events are dependent.

b) Outcomes

In probability theory, an outcome is a possible result of an experiment. Each possible outcome of a particular experiment is unique, and different outcomes are mutually exclusive (only one outcome will occur on each trial of the experiment). All of the possible outcomes of an experiment form the elements of a sample space.

Whenever we do an experiment like flipping a coin or rolling a die, we get an outcome. For example, if we flip a coin we get an outcome of heads or tails, and if we roll a die we get an outcome of 1, 2, 3, 4, 5, or 6. Unless we're rolling a 20-sided die, in which case we're likely playing Dungeons & Dragons, and the outcome is that we won't go on a date for a few years yet.

We call the set of all possible outcomes of an experiment the sample space. The sample space for the experiment of flipping a coin is: {heads, tails} and the sample space for the experiment of rolling a die is {1, 2, 3, 4, 5, 6}.

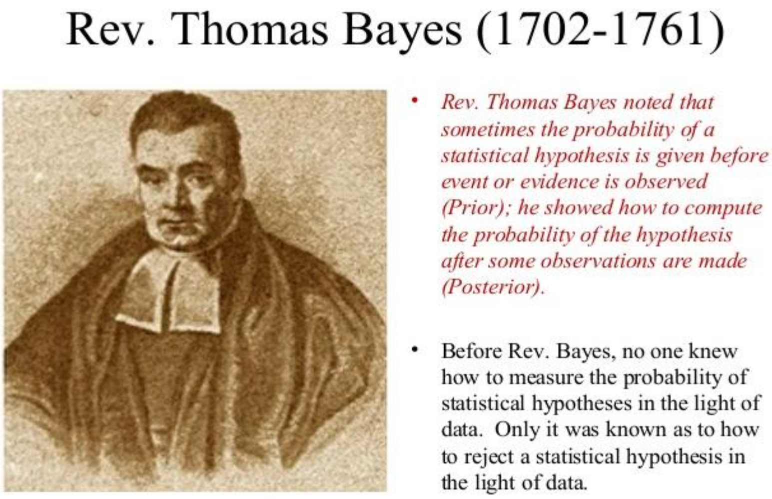
When we talk about finding probabilities, we mean finding the likelihood of events. If an experiment is random/fair, the probability of an event is the number of favorable outcomes divided by the total number of possible outcomes:

No. of favorable outcomes/No. of possible outcomes

A favorable outcome is any outcome in the event whose probability you're finding (remember, an event is a set).

1.6 Conditional probability

Let A and B be two events. Then, the conditional probability of A given that B has occurred, P(A | B), is defined as: $P(A \mid B) = P(A \cap B) / P(B)$



The reasoning behind this definition is that if B has occurred, then only the "portion" of A that is contained in B, i.e., A ∩ B, could occur; moreover, the original probability of A ∩ B must be recalculated to reflect the fact that the "new" sample space is B. This conditional probability is also called as Bayes Rule.

Example 1: Assume once more that you are a doctor. Again, this is the sample space of "patients you might see on any given day". Consider an Event: Flu (F), its probability P(F) = 0.02 and another Event: Headache (H) with probability P(H) = 0.10 as shown in below figure:

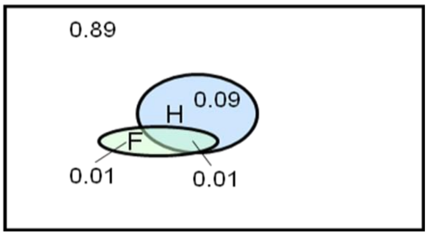


Figure 2. Diagrammatic representation of sample scenario and its probabilities

Suppose we want analyse the interactions between these two events , which may be defined as: **P(H|F)** = Fraction of F's outcomes which are also in H P(H|F)

= Fraction of flu in which patient has a headache

= # with flu and headache / # with flu

= Size of "H and F" region / Size of "F" region

= P(H, F) / P(F)

One day you wake up with a headache. You think: "50% of flus are associated with headaches so I must have a 50-50 chance of coming down with flu", i.e., (P(H|F) = 0.50). Then,

$$P(F \mid H) = \frac{P(F,H)}{P(H)} = \frac{P(H \mid F)P(F)}{P(H)} = \frac{(0.50)(0.02)}{0.1} = 0.10$$

Example 2: Pick a Card from a Deck Suppose a card is drawn randomly from a deck and found to be an Ace. What is the conditional probability for this card to be Spade Ace?

A = Spade Ace
B = an Ace
A $\cap$ B = Spade Ace
P(A) = 1/52; P(B) = 4/52; and P A $\cap$ B = 1/52

Hence, P(A | B) =(1/52) /(4/52)= ¼

1.7 Probabilistic inference

Probabilistic inference is the task of deriving the probability of one or more random variables taking a specific value or set of values. For example, a Bernoulli (Boolean) random variable may describe the event that John has cancer. Such a variable could take a value of 1 (John has cancer) or 0 (John does not have cancer). DeepDive uses probabilistic inference to estimate the probability that the random variable takes value 1: a probability of 0.78 would mean that John is 78% likely to have cancer.

The needs for probabilistic inference are:

1. Probabilistic learning: Calculate explicit probabilities for hypothesis, among the most practical approaches to certain types of learning problems
2. Incremental: Each training example can incrementally increase/decrease the probability that a hypothesis is correct. Prior knowledge can be combined with observed data.
3. Probabilistic prediction: Predict multiple hypotheses, weighted by their probabilities
4. Standard: Even when Bayesian methods are computationally intractable, they can provide a standard of optimal decision making against which other methods can be measured

1.7.1 Application of Bayesian theory

1.7.2 Bayesian Classification

Given training data D, posteriori probability of a hypothesis h, P(h|D) follows the Bayes theorem:

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

MAP (maximum posteriori) hypothesis is defined as follows:

$$h_{MAP} \equiv \underset{h \in H}{\arg\max}\, P(h|D) = \underset{h \in H}{\arg\max}\, P(D|h)P(h).$$

Practical difficulty is that it requires initial knowledge of many probabilities, which introduces significant computational cost.

The classification problem may be formalized using a-posteriori probabilities:

P(C|X) = prob. that the sample tuple X=<x_1,…,x_k> is of class C.

Idea is to assign to sample X the class label C such that P(C|X) is maximal.

Estimating a-posteriori probabilities

As per Bayes theorem:

P(C|X) = P(X|C)·P(C) / P(X)

where, P(X) is constant for all classes

P(C) = relative freq of class C samples

C such that P(C|X) is maximum = C such that P(X|C)·P(C) is maximum

Problem: computing P(X|C) is unfeasible!

Example – Training

Lets say we have a table that decided if we should play tennis under certain circumstances. These could be the **outlook** of the weather; the **temperature**; **the** humidity **and the strength of the** wind:

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

So here we have 4 attributes. What we need to do is to create “look-up tables” for each of these attributes, and write in the probability that a game of tennis will be played based on this attribute. In these tables we have to note that there are 5 cases of not being able to play a game, and 9 cases of being able to play a game.

P(Play=Yes) = 9/14
P(Play=No) = 5/14

We also must note the following probabilities for P(C):

HUMIDITY	Play = Yes	Play = No	Total
High	3/9	4/5	7/14
Normal	6/9	1/5	7/14

WIND	Play = Yes	Play = No	Total
Strong	3/9	3/5	6/14
Weak	6/9	2/5	8/14

OUTLOOK	Play = Yes	Play = No	Total
Sunny	2/9	3/5	5/14
Overcast	4/9	0/5	4/14
Rain	3/9	2/5	5/14

TEMPERATURE	Play = Yes	Play = No	Total
Hot	2/9	2/5	4/14
Mild	4/9	2/5	6/14
Cool	3/9	1/5	4/14

Testing

Now, we are in the testing phase. For this, say we were given a new instance, and we want to know if we can play a game or not, then we need to lookup the results from the tables above. So, this new instance is:

X = (Outlook=Sunny, Temperature=Cool, Humidity=High, Wind=Strong)

Firstly we look at the probability that we can play the game, so we use the lookup tables to get:

P(Outlook=Sunny | Play=Yes) = 2/9
P(Temperature=Cool | Play=Yes) = 3/9
P(Humidity=High | Play=Yes) = 3/9
P(Wind=Strong | Play=Yes) = 3/9
P(Play=Yes) = 9/14

Next we consider the fact that we cannot play a game:
P(Outlook=Sunny | Play=No) = 3/5
P(Temperature=Cool | Play=No) = 1/5
P(Humidity=High | Play=No) = 4/5
P(Wind=Strong | Play=No) = 3/5
P(Play=No) = 5/14

Then, using those results, you have to multiple the whole lot together. So you multiple all the probabilities for Play=Yes such as:
P(X|Play=Yes)P(Play=Yes) = (2/9) * (3/9) * (3/9) * (3/9) * (9/14) = 0.0053
And this gives us a value that represents 'P(X|C)P(C)', or in this case 'P(X|Play=Yes)P(Play=Yes)'.

We also have to do the same thing for Play=No:
P(X|Play=No)P(Play=No) = (3/5) * (1/5) * (4/5) * (3/5) * (5/14) = 0.0206

Finally, we have to divide both results by the evidence, or 'P(X)'. The evidence for both equations is the same, and we can find the values we need within the 'Total' columns of the look-up tables. Therefore:
P(X) = P(Outlook=Sunny) * P(Temperature=Cool) * P(Humidity=High) * P(Wind=Strong)
P(X) = (5/14) * (4/14) * (7/14) * (6/14)
P(X) = 0.02186

Then, dividing the results by this value:
P(Play=Yes | X) = 0.0053/0.02186 = 0.2424
P(Play=No | X) = 0.0206/0.02186 = 0.9421

So, given the probabilities, can we play a game or not? To do this, we look at both probabilities and see which once has the highest value, and that is our answer. Therefore:
P(Play=Yes | X) = 0.2424
P(Play=No | X) = 0.9421

Since 0.9421 is greater than 0.2424 then the answer is 'no', we cannot play a game of tennis today.

Last few points about Naïve Bayes is that since it isbased on the independence assumption:

- Training is very easy and fast; just requiring considering each attribute in each class separately
- Test is straightforward; just looking up tables or calculating conditional probabilities with normal distributions

Summary

- Probability theory is very essential for analytics.
- Inferences based on probability theory are effective.
- Analytics based on probability theory are widely used.