Learning Probability and Statistics for Machine Learning and Data Science

Parth Sinha
Student of AI & ML
Oxford Brookes University
Email: sinhaparth555@gmail.com

©Parth Kumar Sinha 2024

***Abstract*—This is a note for Probability and Statistics for Machine Learning and Data Science.**

***Index Terms*—data science, machine learning, probability and statistics**

I. Probability

A. Introduction

Probability is the measure of the likelihood or chances that an event will occur.
For example, person A and person B are working in a company, so what are the changes for person A who is wealthier than person B. So we need to quantify how much person A earns more than person B? This quantification is known as probability.
The probability range between

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

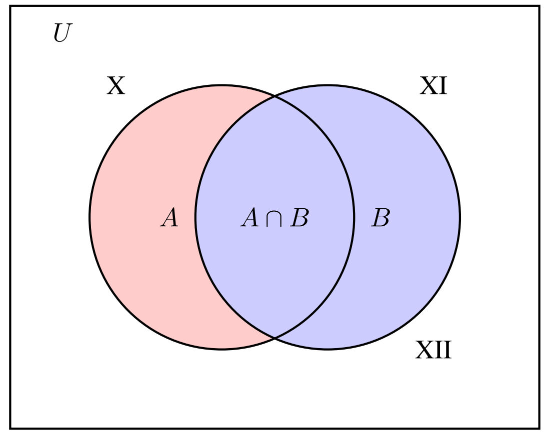
where 0 means that the probability of the event occurring is null and 1 means that the probability of the event occurring is true. For example $P(sun\ will\ rise\ for\ west) = 0$ and similarly $P(sun\ will\ rise\ from\ east) = 1$.
Sample Space: The sample space of an experiment or random trial is the set of all possible outcomes or results of that experiment.
Example:
Experiment: Rolling a standard six-sided die.
Sample Space: $\{1, 2, 3, 4, 5, 6\}$
This means that the possible outcomes are rolling a 1, 2, 3, 4, 5, or 6.
The sample space is the set of all possible outcomes.
Case $\Rightarrow$ Odd no: $\{1, 3, 5\}$
An event is something that happens, and event is always a subset of sample space. $E \subseteq S$ where E = event and S = sample space.
For example

$$E = \{1, 3, 5\} \Rightarrow n(E) = 3$$

$$S = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(S) = 6$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$$

If the dice is rolled fairly 100 times, 50% times it will give odd numbers.



U: All Students
A: Class IX
B: Computer Science

Fig. 1. Venn diagram showing relationship between Class IX students and Computer Science students

B. Sets for probability

1) Intersection: Students who are in class IX and the students who have opted for computer science $\Rightarrow n(A \cap B)$

2) Union: Students who are in class IX and the students who are in computer science $\Rightarrow n(A \cup B)$

$$n(A \cap B) = n(A) + n(B) - n(A \cap B)$$

3) Null Sets: Class XI students opted Economics and also Class XI Science batch student have opted for Economics $n(A \cap B) = \emptyset$

C. Fair Events

Fair events in probability are events where each possible outcome has an equal chance of occurring. This means that there is no bias or favoritism towards any particular outcome. For example: Coin in which tails and heads equally likely to occurs $\Rightarrow P(H) = 0.5$ where H is head of the coin.
Take a fair coin and toss n times.

In table I when $n \to \infty$ the probability is head is 50% i.e., $P(H) \to 0.5$.

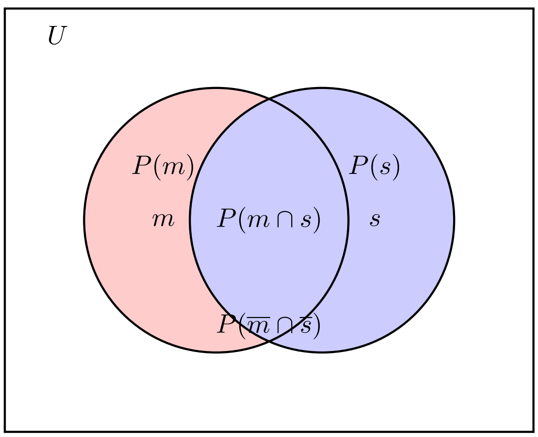
©Parth Kumar Sinha 2024

TABLE I
OCCURRENCE OF OF HEADS n TIMES

n	Occurrence of head
10	6
100	52
1000	512
.	.
.	.
.	.
1 million	500010

D. *Joint Probability*

Joint probability is a fundamental concept in probability theory that measures the likelihood of two or more events occurring simultaneously. It's often represented as $P(A \cap B)$, where A and B are the events in question.
For example, U is the total malaria patient, m no. of patient getting medicine and s is no. of patient survive, refer to figure 2. So, $P(m \cap s) \to$ no. of malaria patient those are given medicine and also survived.



U: Total patients
m: Patients getting medicine
s: Patients who survive

Fig. 2. Joint probability distribution of malaria patients receiving medicine and survival

For example
So, if total malaria patient are 500, 50 of them got medicine and 200 survived, and patient who got medicine and survived are 25, then:

$$U = 500$$

$$m = 50$$

$$s = 200$$

$$m \cap s = 25$$

$$\text{Then, } P(m \cap s) = \frac{25}{500} = \frac{1}{20} = 0.05$$

Question: Fair dice throw:

A : *even no.*
B : *prime no.*
C : *odd no.*

Ans *a)* $P(A \cap B \cap C)$

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

$$A : \{2, 4, 6\}$$

$$B : \{2, 3, 5\}$$

$$C : \{1, 3, 5\}$$

$$n(A \cap B \cap C) = 0$$

$$P(A \cap B \cap C) = \frac{0}{6} = 0 = \emptyset$$

Ans *b)* $P(even\ or\ prime)$

$$A : \{2, 4, 6\}$$

$$B : \{2, 3, 5\}$$

$$n(A \cup B) = n(5)$$

$$P(A \cup B) = \frac{n(A \cup B)}{n(U)} = \frac{5}{6} = 0.83$$

E. *Conditional Probability*

Conditional probability is a measure of the probability of an event occurring, given that another event has already occurred.[1] It essentially refines our understanding of probability by considering the impact of prior knowledge or conditions.
Equation,

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

For s is survived and m is medicine given to patient then, $P(s|m) \to$ Probability that the patient survives, given that he's been administrated the medicine. Then,

$$s = 200$$

$$m = 50$$

$$m \cap s = 25$$

$$U = 500$$

$$P(s|m) = \frac{P(s \cap m)}{P(m)}$$

$$\frac{\frac{n(s \cap m)}{n(U)}}{\frac{n(m)}{n(U)}}$$

$$\frac{n(s \cap m)}{n(m)} = \frac{25}{50} = \frac{1}{2} = 0.5$$

F. Types of events

1) Mutually Exclusive events: two events are considered mutually exclusive if they cannot both occur simultaneously. This means that if one event happens, the other event cannot happen.
Fair dice is throw:
$A : no.\ is\ even = \{2, 4, 6\}$
$B : no.\ is\ odd = \{1, 3, 5\}$

$$A \cap B = 0 = \emptyset$$

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

2) Independent events: In probability theory, two events are considered independent if the outcome of one event does not affect the probability of the other event occurring.[2]
For example:
Total cards: 52
$A : card\ is\ a\ king$
$B : Card\ is\ a\ heart$

$$P(A) = \frac{4}{52} = \frac{1}{13}$$

$$P(B) = \frac{13}{52} = \frac{1}{4}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{52}}{\frac{1}{4}} = \frac{4}{52} = \frac{1}{13}$$

$$P(A|B) = P(A)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{52}}{\frac{1}{13}} = \frac{13}{52} = \frac{1}{4}$$

$$P(B|A) = P(B)$$

If the occurrence of one event does not alter the changes of occurrence of another event knows as independent events.

$$P(E_1|E_2) = P(E_1)$$

$$P(E_2|E_1) = P(E_2)$$

G. Co-joint Probability

Joint probability is a fundamental concept in probability theory that measures the likelihood of two or more events occurring simultaneously.
Equation,

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

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

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

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

Question: n-coins toss, probability all heads?
$Sol^n : E_i : coin\ is\ head$

$$P(E_1 \cap E_2 \cap E_3 ...) = P(E_1) \times P(E_2) \times P(E_3)...$$

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}...$$

$$= \left(\frac{1}{2}\right)^n$$

Lets see,

$$P(E_1 \cap E_2 \cap E_3) = P(E_1 \cap (E_2 \cap E_3))$$

$$= P(E_1) \cdot P(E_2 \cap E_3)$$

$$= P(E_1) \cdot P(E_2) \cdot P(E_3)$$

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

$$P(E_1) \cdot P(E_2|E_1) = P(E_1 \cap E_2)$$

$$P(E_1)P(E_2) = P(E_1 \cap E_2)$$

H. Bayes' Theorem:

Bayes' Theorem is a mathematical formula that helps you update your beliefs (or probabilities) about something based on new evidence.[3] It allows you to go from a prior belief to a posterior belief after observing new data.[4]

Bayes' Theorem problem from the book Think Bayes'[5]

The Cookie problem

Suppose there are two bowls of cookies. Bowl 1 contains 30 vanilla cookies and 10 chocolate cookies. Bowl 2 contains 20 of each.
Now suppose you choose one of the bowls at random and, without looking select a cookie at random. The cookie is vanilla. What is the probability that it came from Bowl 1?
This is a conditional probability, we want $p(Bowl\ 1|vanilla)$, but it is not obvious how to compute it. If I asked a different question - the probability of a vanilla cookie given Bowl 1 - it would be easy:

$$p(vanilla|Bowl\ 1) = \frac{3}{4}$$

Sadly, $p(A|B)$ is *not* the same as $p(B|A)$, but there is a way to get from one to the other: Bayes' theorem.

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

Proof,

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

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

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

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

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

$$P(A|B) \cdot P(B) = P(A) \cdot P(B|A)$$

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

©Parth Kumar Sinha 2024

TABLE II
TRUE & FALSE TABLE

Question 1	Question 2	Question 3
True	True	True
True	True	False
True	False	True
True	False	False
False	True	True
False	True	False
False	False	True
False	False	False

Cookie problem continue:

$$p(Bowl\ 1|vanilla) = \frac{p(Bowl\ 1) \cdot p(vanilla|Bowl\ 1)}{p(vanilla)}$$
$$= \frac{\frac{1}{2} \cdot \frac{3}{4}}{\frac{50}{80}}$$
$$= \frac{3}{5}$$
$$P(Bowl\ 1|vanilla) = \frac{P(Bowl\ 1 \cap vanilla)}{P(vanilla)}$$

Why we use Bayes' theorem in Data Science and Machine Learning
Let say we want to know what are the chances of getting heart attack to a person, we will note down some conditions like, cholesterol, BMI, etc,.
Lets say $P(H) \Rightarrow P(H|Ch \uparrow)$ the probability of the heart attack will be high and if the $P(H|Ch \downarrow)$ the probability of the heart attack will be low, Similarly with BMI.

I. Combinatorics

Question $\Rightarrow$ Given that we have 3 questions of True and False, in how many ways we can answer them?
Sol^n: Each question has 2 possible answers (True or False). Since there are 3 questions, the total number of ways to answer them is $2 \times 2 \times 2 = 2^3 = 8$. Please see the table II.

J. Permutations

Permutations in mathematics refer to the different arrangements of a set of objects where the order of the objects matters. For example $(i, j) \neq (j, i)$, ex. Student 1 got rank 1 and student 2 is got rank 2 but it's not possible that student 2 get rank 1 and student 1 get rank 2.

$$(S_1, S_2) \neq (S_2, S_1)$$

Question $\Rightarrow$ There are 3 alphabets (a, b, c), how many ways to arrange them?
$Sol^n \Rightarrow$ There are 6 ways to arrange the alphabets (a, b, c), see table III.

Formula for factorial:

$$n! = n \times (n-1) \times (n-2) \times ... \times 1$$

Lets take an example, there are n items and we have to arrange them in r places. So lets take a look:

TABLE III
WAYS TO ARRANGE APLPHABETS (a, b, c)

Position 1	Position 2	Position 3
a	b	c
a	c	b
b	a	c
b	c	a
c	a	b
c	b	a

TABLE IV
STUDENTS COMBINATIONS

Combinations	*Student 1*	*Student 2*	*Student 3*
1	S_1	S_2	S_3
2	S_1	S_2	S_4
3	S_1	S_3	S_4
4	S_2	S_3	S_4

Lets say we have 5 objects i.e., (a, b, c, d, e) and we have to arrange them in 2 places, $n = 5$ and $r = 2$, Then using the formula we get $5 \times (5-1) = 5 \times 4 = 20$.
Lets say we have N objects and we need to arrange them in r places.
$(n)\ and\ (n-1)\ and\ (n-2)...(n-r+1)$
No. of ways $= n \times (n-1) \times (n-2)...(n-r+1)$

$$= \frac{(n \times (n-1) \times (n-2)...(n-r+1)) \times ((n-r) \times (n-r-1)...1}{((n-r) \times (n-r-1) \times ... \times 1}$$
$$\Rightarrow \frac{n!}{(n-r)!}$$
$$^nP_r = \frac{n!}{(n-r)!}$$

K. Combinations

Combinations in mathematics refer to the selection of a subset of items from a larger set, where the order of selection does not matter. The section of objects are equal i.e., $(i, j) = (j, i)$. For example, if we have to select student S_1 or S_2 in any combination it's equal, i.e., $(S_1, S_2) = (S_2, S_1)$
Lets take little complicated example, we need to select 3 students out of 4 students, and we represent all students with S_1, S_2, S_3 and S_4.
Sol^n: Total no. of arrangements = 24 and no. of selections = $\frac{24}{3!}$
$= 4$
You can also see the table IV.

No. of arrangements $\Rightarrow {}^nP_r = \frac{n!}{(n-r)!}$

$$\frac{\frac{n!}{(n-r)!}}{r!}$$

No. of selections = $\frac{n!}{(n-r)! \cdot r!}$

$$^nC_r = \frac{n!}{(n-r)! \cdot r!}$$

©Parth Kumar Sinha 2024

Lets take an example, we have set of alphabets $\{a, b, c, d, e\}$ and we want to select only 3 of them, lets solve it with the equation.

$$^nC_r = \frac{n!}{(n-r)!r!} \Rightarrow \frac{5!}{(5-3)!3!} = 10$$

So we can have 10 selections.

Question $\rightarrow$ $^nC_0 +^n C_1 +^n C_2 + ... +^n C_n =?$
$^nC_r \rightarrow no.\ of\ selections$
$Sol^n \Rightarrow \{1, 2, 3\}$
Lets say I don't want to select any one from the set, so $\rightarrow r = 1$ and so on, so we have now

$$^nC_0 = 1 = \{\}$$

$$^nC_1 = \{1\},\ \{2\},\ \{3\}$$

$$^nC_2 = \{1, 2\},\ \{1, 3\},\ \{2, 3\}$$

$$^nC_3 = \{1, 2, 3\}$$

$S = \{1, 2, 3\}$
$\{1\},\ \{2\}...$
$^nC_0 \rightarrow^n C_n$
$S = 2^n$
$\{a_1,\ a_2,\ a_3, ...,\ a_n\}$
So basically solving this we get:

$$^nC_0 +^n C_1 +^n C_2 + ... +^n C_n = 2^n$$

Question $\rightarrow$ $^nC_n =^n C_{n-r}$

$$\frac{n!}{(n-r)!r!} = \frac{n!}{(n-(n-r))!(n-r)!}$$

$$= \frac{n!}{n!(n-r)!}$$

Then above equation doesn't work
Let $5 \rightarrow n$ and we need to select only 2 i.e., $2 \rightarrow r$ so it will become $^5C_2 = x$, where x is the result.
Now again, let's say $5 \rightarrow n$ and we need to removed 3 i.e., $3 \rightarrow r$ so it will become $^5C_3$, let take set $\{P_1,\ P_2,\ P_3,\ P_4,\ P_5\}$, now you can take a look at table V.
So as we look at table V we can say that

$$^3C_2 =^5 C_3$$

so we get

$$^nC_r =^n C_{n-r}$$

P_1P_2	$P_3P_4P_5$
P_2P_3	$P_2P_4P_5$
P_1P_4	$P_2P_3P_3$
P_1P_3	$P_2P_4P_3$

TABLE V
SELECTION AND REMOVING ELEMENTS

II. Probability Distribution

A probability distribution is a mathematical function that describes the likelihood of different possible values for a random variable within a given range.[6]

A. *Random variables*

A random variable is a function that assigns a numerical value to each possible outcome of a random experiment.[7]

For example, if we roll the dice i.e., $S = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ and every number is countable, another we assign grades to students i.e.m $S = \{A,\ B,\ C,\ D,\ E,\ F\}$

1) *Discrete:* Set of countable.
2) *Continuous:* Set of all possible values within an interval.
3) *Ordinal:* Type of statistical data where the variables have a defined order or ranking.

Mean: the mean is the average of a set of numbers.
Let say we take 5 students age, and we have 10, 11, 12, 9 and 8, so we can get the mean of the student's age.

$$Mean: \frac{10 + 11 + 12 + 9 + 8}{5} = \frac{50}{5} = 10$$

So the mean of the students age is 10.

References

[1] N. Rosidi, "Machine learning algorithms you should know for data science." [Online]. Available: https://www.stratascratch.com/blog/machine-learning-algorithms-you-should-know-for-data-science/
[2] C. S. R. Center, "Statistically independent events." [Online]. Available: https://csrc.nist.gov/glossary/term/statistically_independent_events#:~:text=Definitions%3A,occurrence%20of%20the%20other%20event.
[3] S. Taylor, "Bayes' theorem," Mar 2023. [Online]. Available: https://corporatefinanceinstitute.com/resources/data-science/bayes-theorem/
[4] M. Taboga, "Bayesian inference — introduction with explained examples," 2021. [Online]. Available: https://www.statlect.com/fundamentals-of-statistics/Bayesian-inference
[5] A. B. Downey, *Think Bayes.* " O'Reilly Media, Inc.", 2021.
[6] NCL, "Numeracy, maths and statistics - academic skills kit." [Online]. Available: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/distribution-functions/probability-density-function.html
[7] Numerade, "show from the basic definition of random variable that a maxxy is a random variable and that b minxy max x y is a random variable 49846," 2023. [Online]. Available: https://www.numerade.com/ask/question/show-from-the-basic-definition-of-random-variable-that-a-maxxy-is-a-random-variable

©Parth Kumar Sinha 2024