Lab 5 Conditional Probability

In probability theory, conditional probability is the probability of occurring of an event where it is given that another event has already occurred. Try to ***extend this idea of calculating conditional probabilities to Bayes Theorem*** by taking the below dataset.

https://www.kaggle.com/mukulthakur177/flood-prediction-model?select=kerala.csv

The dataset contains the monthly rainfall data from years 1901 to 2018 for the Indian state of Kerala.

1. Calculate the following

   - Calculate probability of flood given it rained more than 500 mm in June
   - Calculate Probability of rain more than 500 mm in June given it flooded that year
   - Calculate probability of flood given it rained more than 500 mm in July
   - Calculate Probability of rain more than 500 mm in July given it flooded that year

Explanation:

Events and Sample Space

Event is just the result of a random experiment.

Ex: getting a head when we toss a coin is one event

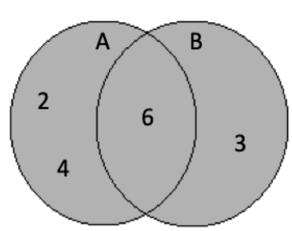
A collection of all possible outcomes of an event is called a sample space

Ex: For tossing the coin we can have just two outcomes: head (H) or a tail(T).

Ex: rolling a fair die will always result in some number between 1 to 6, hence the sample space is {1, 2, 3, 4, 5, 6}.

Union of events

- **Event A:** Getting a number which is divisible by 2
- **Event B:** Getting a number which is divisible by 3
- The sample space for event A is {2, 4, 6} whereas for event B it is {3, 6}.
- Now if we define another event C which is getting a number which is divisible by either 2 **OR** 3 our new sample space would just be the combination of all the unique elements of sample space A and sample space B: {2, 3, 4, 6}.
- Mathematically, these events can be shown in terms of venn diagrams:

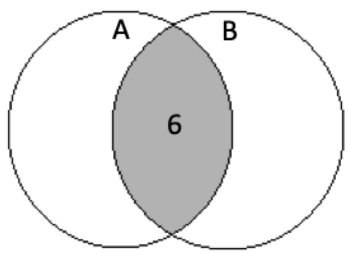


In terms of probabilities, we can easily calculate the probabilities for all the events as follows:

- P(A) = $\frac{\text{Number of cases where the output is divisible by 2}}{\text{Total possible outcomes}}$ = 3/6 = 0.5
- P(B) = $\frac{\text{Number of cases where the output is divisible by 3}}{\text{Total possible outcomes}}$ = 2/6 = 0.333
- P(C) = P(A ∪ B) = $\frac{\text{Number of cases where the output is divisible by either 2 OR 3}}{\text{Total possible outcomes}}$ = 4/6 = 0.667

Intersection of events

Following the events defined previously, we can also define an event D which is getting a number which is divisible by both 2 **AND** 3, meaning *the common element in the sample space of both the events*. In terms of venn diagrams, it can be shown as:



Again, in terms of probabilities:

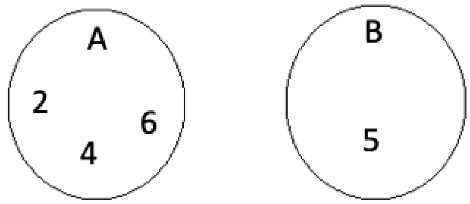
- P(A) = $\frac{\text{Number of cases where the output is divisible by 2}}{\text{Total possible outcomes}}$ = 3/6 = 0.5
- P(B) = $\frac{\text{Number of cases where the output is divisible by 3}}{\text{Total possible outcomes}}$ = 2/6 = 0.333
- P(C) = P(A ∩ B) = $\frac{\text{Number of cases where the output is divisible by both 2 AND 3}}{\text{Total possible outcomes}}$ = 1/6 = 0.167

Disjoint events

Consider these two events:

- **Event A:** Getting a number which is divisible by 2
- **Event B:** Getting a number which is divisible by 5
- The sample space for event A is {2, 4, 6} whereas for event B it is {5}.
- We can see that these events cannot occur together in the case of rolling a fair die.
- These events are called disjoint events.

- The venn diagram for these type of events can be shown as:



Dependent and Independent events

If the occurrence of one event does not effect the occurrence of another event, then these events are termed as independent events. Few examples of independent events include:

- Getting a head when a coin is tossed AND getting a 5 in rolling a fair die
- Getting rains in the month of August AND snow in December

The probability in the case of independent events can be written as P(A∩∩B) = P(A) * P(B), that is, the probability of occurring both the events is just the product of individual probabilities.

Conditional probability can be defined as follows:

Probability of event A given event B has already occurred = P(A|B) = $\frac{P(A \cap B)}{P(B)}$

we can easily see the above equation reduces to `P(A)` for independent events by writing `P(A ∩ B) = P(A)*P(B)`.

```
# Import libraries
import numpy as np
import pandas as pd
```

```
# Read the data
df = pd.read_csv("/kaggle/input/kerela-flood/kerala.csv")
df.head()
```

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
0	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5
1	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8
2	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2
3	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8
4	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6

Defining some variables:

- `P(F)` : Probability of flooding
- `P(J)` : Probability of having more than 500 mm rain in June
- `P(F ∩ J)` : Probability of flooding **and** having more than 500 mm rain in June
- `P(F|J)` : Probability of flooding given it rained more than 500 mm in June

Based on the above table we can easily find these probabilities.

```
P_F = (6 + 54) / (6 + 54 + 19 + 39)
P_J = (39 + 54) / (6 + 54 + 19 + 39)
P_F_intersect_J = 54 / (6 + 54 + 19 + 39)
print(f"P(F): {P_F}")
print(f"P(J): {P_J}")
print(f"P(F AND J): {P_F_intersect_J}")
```

```
P(F): 0.5084745762711864
P(J): 0.788135593220339
P(F AND J): 0.4576271186440678
```

Using the formula - `P(A|B) = P(A ∩ B) / P(B)` we can easily calculate the conditional probability:

```
# Now calculate probailitity of flood given it rained more than 500 m
m in June (P(A|B))
P_F_J = P_F_intersect_J / P_J
print(f"P(F|J): {P_F_J}")
```

```
P(F|J): 0.5806451612903226
```

given that it flooded in Kerala in a given year what is the probability that it rained more than 500 mm in the month of June or July?

Here **Bayes Theorem** comes into action. Some other examples of Bayes Theorem are like:

- The probability of a woman having breast cancer given she tested positive in the test
- Probability that a given email is actually a spam given it contains certain flagged words.

Bayes Theorem can be easily derived using the relationship between conditional probability and intersection of events. Given two events, we already know:

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

$$\text{so, } P(B|A) = \frac{P(A|B).P(B)}{P(A)}$$

In Bayesian inference, `P(B)` is called **Prior Probability**. In our case, `P(J)` is the prior probability which tells the probability of rain more than 500 mm in June (or July) without knowing whether it flooded or not that year. We can see prior probability is the probability of the event we are interested in before any new information.

```
# Probability of rain more than 500 mm in June given it flooded that
year (P(B|A))
P_J_F = (P_F_J * P_J) / P_F
print(f"P(J|F): {P_J_F}")
```

```
P(J|F): 0.9000000000000001
```

Defining the similar parameters for July:

- `P(F)` : Probability of flooding
- `P(J)` : Probability of having more than 500 mm rain in July
- `P(F ∩ J)` : Probability of flooding **and** having more than 500 mm rain in July
- `P(F|J)` : Probability of flooding given it rained more than 500 mm in July

```
P_F = (3 + 57) / (3 + 57 + 19 + 39)
P_J = (39 + 57) / (3 + 57 + 19 + 39)
P_F_intersect_J = 57 / (3 + 57 + 19 + 39)
print(f"P(F): {P_F}")
print(f"P(J): {P_J}")
print(f"P(F AND J): {P_F_intersect_J}")
```

```
P(F): 0.5084745762711864
P(J): 0.8135593220338984
P(F AND J): 0.4830508474576271
```

```python
# Now calculate probailitity of flood given it rained more than 500 mm in July
P_F_J = P_F_intersect_J / P_J
print(f"P(F|J): {P_F_J}")
```

```
P(F|J): 0.59375
```

```python
# Probability of rain more than 500 mm in July given it flooded that year (P(B|A))
P_J_F = (P_F_J * P_J) / P_F
print(f"P(J|F): {P_J_F}")
```

```
P(J|F): 0.9500000000000002
```