STU11002

Statistical Analysis I

Learning objectives

- Define the terms probability, experiment, and event.
- Understand Venn Diagram operations and their associated symbols in the context of probability.
- Create estimates for the probability of events when given the relevant information.
- State and use the Chain rule for conditional probabilities.
- Define independence in the context of probability.
- Define and use Bayes Theorem for conditional probabilities.

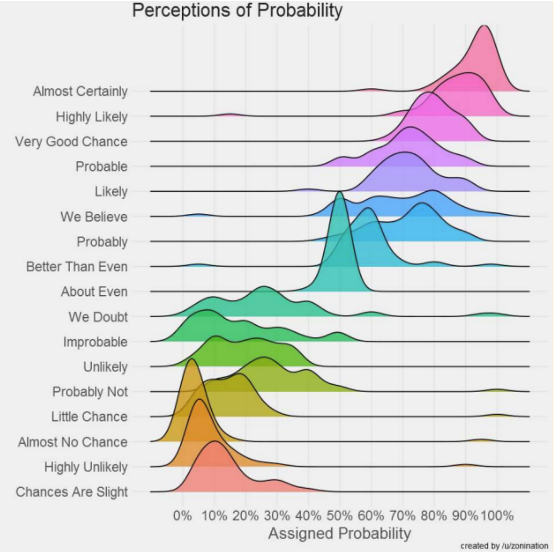
Probability

- An introduction -

What is probability?

- **Uncertainty** is all around us, in everyday life
- The study of **probability** is the study of **randomness**
- We can trace back the origin of probability to 1600, with Blaise Pascal and Pierre de Fermat
- Goal of probability: How can mathematics be used to describe random events?

How do we quantify randomness?



Building blocks

Probability theory is based on 3 concepts:

- **Experiment**
- **Event**
- **Probability**

Experiment: An experiment is any procedure that can be replicated an infinite number of times and for which a defined set of possible outcomes is available. Such set of possible outcomes is called the **sample space**. The sample space contains two or more possible outcomes. If there is only one possible outcome, then the experiment is said to be **deterministic**. Multiple repetitions of the same experiment are called **trials**.

Building blocks

Event: An event is a set of outcomes from an experiment. Each event happens with probability $Pr(E)$, where E denotes the event. An event that has a single outcome is called **elementary**, while an event with more outcomes is called **compound**.

Example

Let us consider the experiment "roll of a standard dice" (with 6 faces). The probability of the elementary event $\{1\}$ is $\frac{1}{6}$, while that of the compound event "even number" is $\frac{3}{6} = \frac{1}{2}$.

Building blocks

Probability: Probability is a number comprised between 0 and 1, which measures the degree of uncertainty associated to the realization of an event. When $Pr(E) = 1$, we say that the event is **certain**, while when $Pr(E) = 0$, we say that the event is **impossible**.

Example

Let us consider the experiment "roll of a standard dice" (with 6 faces). The elementary events here are: $\{1; 2; 3; 4; 5; 6\}$. An example of compound event is "number larger than 3": $\{4; 5; 6\}$, or "even number": $\{2; 4; 6\}$.

Some definitions

- We denote with Ω the sample space.
- We denote with $\mathcal{E} = \{E_1, \ldots, E_R\}$ the set of all subsets of events in the sample space Ω. The events in $\mathcal{E}$ have a particular structure (called "Boolean algebra"), which "simply" means that all of the sets of **rules** and **operations** one needs in probability theory are valid for $\mathcal{E}$.

Operations:

There are 3 fundamental operations: **complementation**, **union**, and **intersection**.

Some definitions

Operations

- **Complementation**. The **complement** of a given event A is denoted as $\bar{A}$. This event corresponds to "A does not happen".
- **Union**. The **union** of two events A and B is denoted with $A \cup B$. It means "at least one of the two events A or B happens".
- **Intersection**. The **intersection** of two events A and B is denoted with $A \cap B$. It means "both events A and B happen (jointly)".

Some definitions

Operations - some properties

- $Pr(A) = 1 - Pr(\bar{A})$.
- Distributive property (valid for union and intersection):
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- Impossible event: $A \cap \bar{A} = \emptyset$
- Certain event: $\bar{\emptyset} \equiv \Omega$
- $A \cup \bar{A} = \Omega$
- $\mathcal{E} \subset \Omega$

Some definitions

Operations - some properties

- If Σ contains K elementary events, than $\mathcal{E} = \{E_1, \dots, E_R\}$ is made up of $R = 2^K$ subsets.

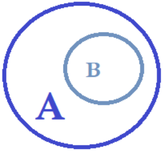
  Example. Assume that Ω contains 3 elementary events: A, B, C. Then we have $2^3 = 8$ possible subsets of events:

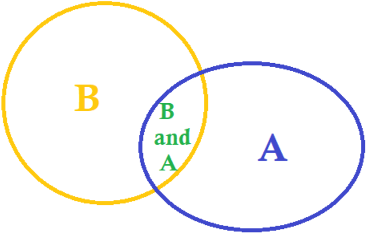
  $$\{\emptyset; A; B; C; (A, B); (A, C); (B, C); (A, B, C) \equiv \Omega\}$$

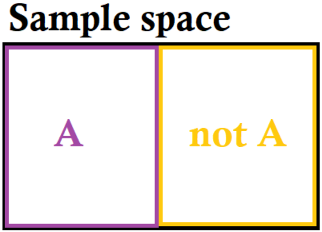
- Events A and B are **incompatible** if $A \cap B = \emptyset$
- If $A \cup B = A$, then $B \subset A$, that is, B is **included** in A

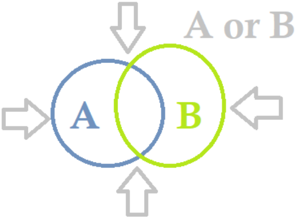
Some definitions

Operations - some properties









Postulates

- Probability is a function that links to every event $E_r \in \mathcal{E}$, $r = 1, \dots, R$, a real number. It is denoted with $Pr(E_r)$.
- Given an event A, $Pr(A) \geq 0$.
- The probability of the certain event is 1: $Pr(\Omega) = 1$.
- If two events are **mutually exclusive**, they can not happen together. The probability of their union is given by the sum of their probabilities:

$$[A \cap B = \emptyset] \implies [Pr(A \cup B) = Pr(A) + Pr(B)]$$

Note: By postulates we mean those assumptions that do not need to be proved.

Postulates

Properties deriving from the postulates

- How to compute the probability of a union:

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

- $0 \leq Pr(A) \leq 1$
- $Pr(\emptyset) = 0$
- $[B \subset A] \implies [Pr(B) \leq Pr(A)]$
- $Pr(\bar{A}) = 1 - Pr(A)$
- $[Pr(A) = 1] \implies [Pr(A \cap B) = Pr(B)]$
- $[Pr(A) = 0] \implies [Pr(A \cup B) = Pr(B)]$
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

Calculating probabilities

According to the "classical" definition, we can compute the **probability of an event** as the ratio between the **number of possible outcomes for the event** and the `total number of possible outcomes`.

Example:

A drawer contains 13 pairs of black socks, 4 pairs of green socks, and 7 pairs of blue socks. What is the probability that, when grabbing one sock from the drawer in the dark, this sock is blue?

$$Pr(\texttt{one blue sock}) = \frac{14}{26 + 8 + 14} = \frac{14}{48} = 0.292$$

NOTE: the classical definition probability assumes that all possible events are equally likely to happen.

Conditional probabilities

In some situations, we may be interested in calculating the probability of an event, `given that` we know that another event has already happened.

Example

Consider a bag containing 18 marbles, 6 red and 12 white. We extract 2 marbles **without replacement**, and we'd like to compute the probability that the first marble is red (event A), while the second marble is also red (event B).

We know how to we can compute $Pr(A) = \frac{6}{18}$. How do we compute $Pr(B)$? This probability will depend on if A happened or did not happen at the first extraction (it is **conditional** on it). We denote this **conditional probability** with: $Pr(B \mid A)$ (probability of B given A).

Conditional probabilities

Example

- If A has happened, prior to the second extraction we have a new sample space, that is the original one, minus the event A.
- In this “new” sample space, we are interested in $Pr(B)$.
- This means that we are actually interested at all possible outcomes with $(A \cap B)$ from the start (as we'd like both B and A to happen).
- When at the second extraction, all possible outcomes that one may consider (as they include $(A \cap B)$ outcomes) are those outcomes that that include A.

Conditional probabilities

From the example, we can then conclude that:

$$Pr(B \mid A) = \frac{Pr(A \cap B)}{Pr(A)}$$

That is, the conditional probability of B given A is found as the ratio between the probability of the intersection of A and B, and the probability of the conditioning event A.

NOTE: $Pr(B \mid A) > 0$ only if $Pr(A) > 0$

Conditional probabilities

- $Pr(A \cap B) = Pr(A)Pr(B \mid A)$
- **Chain rule**:
  $Pr(A \cap B) = Pr(A)Pr(B \mid A) = Pr(B)Pr(A \mid B)$
- $Pr(A \cap B \cap C) = Pr(A)Pr(B \mid A)Pr(C \mid B \cap A)$
- **Independence**. Two event are said to be **independent** if the happening of one has no impact on the probability of the other one, and vice versa:

  $$Pr(A \mid B) = Pr(A) \quad \text{and} \quad Pr(B \mid A) = Pr(B)$$

  From this follows the fact that the probability of the intersection between two independent events is given by the product of their probabilities:

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

Independent events

Two events are then said to be **independent** if and only if:

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

Example

Consider a standard dice. What is the probability that, when rolled, the outcome will be either 2 or 3 (event A), and that an even number will be the outcome (event B). We know that:

$$Pr(A) = \frac{2}{6} = \frac{1}{3}; \quad Pr(B) = \frac{3}{6} = \frac{1}{2}$$

The probability of the intersection of the events A and B is $Pr(A \cap B) = \frac{1}{6}$. Then: $Pr(A \mid B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} = Pr(A)$, and $Pr(B \mid A) = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2} = Pr(B)$. The events are independent!

Example

Sums of two dices

- Total number of outcomes is $6 \times 6 = 36$
- Total number of unique outcomes is 11
- Example: the probability of the sum being equal to 10 is $\frac{3}{36} = \frac{1}{12}$

Value	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Example

Sums of two dices - empirical and theoretical probabilities

Sum	100 Tries	1000 Tries	10000 Tries	Actual value
2	0.03	0.03	0.0289	0.027778
3	0.03	0.051	0.0552	0.055556
4	0.1	0.085	0.0871	0.083333
5	0.1	0.113	0.1148	0.111111
6	0.09	0.135	0.1379	0.138889
7	0.19	0.182	0.1672	0.166667
8	0.19	0.128	0.1359	0.138889
9	0.13	0.109	0.1049	0.111111
10	0.08	0.073	0.0805	0.083333
11	0.04	0.066	0.0598	0.055556
12	0.02	0.028	0.0278	0.027778

Bayes Theorem

Example

From previous studies, we know that the incidence of disease A in the population of interest is 15%. To detect disease A, a blood test is needed. However, such test is associated to a false negative rate of 15% (test says no disease, but the disease is present), and to a false positive rate of 20% (test founds the disease, but the disease is not present).

What is the probability that, if a patient tests positive, then the patient actually has the disease?

Bayes Theorem

Example

What is the probability that, if a patient tests positive, then the patient actually has the disease?

Here, we nee to consider 4 events:

- The patient has the disease (A)
- The patient does not have the disease ($\bar{A}$)
- The test is negative ($\bar{B}$)
- The test is positive (B)

What we are interested in is: $Pr(A \mid B)$

Bayes Theorem

Derivation

- Consider a set of events $\{A_1, \ldots, A_i, \ldots, A_j, \ldots, A_K\}$, such that $A_i \cap A_j = \emptyset$, for all $i, j = 1, \ldots, K$. Also, $\cup_{i=1}^{K} A_i = \Omega$.
- Consider another event $B \subset \Omega$.
- Then...

$$
\begin{aligned}
Pr(B) = Pr(B \cap \Omega) &= Pr\left(B \cap \cup_{i=1}^{K} A_i\right) \\
&= Pr\left([A_1 \cap B] \cup [A_2 \cap B] \cup \cdots \cup [A_K \cap B]\right) \\
&= Pr(A_1 \cap B) + Pr(A_2 \cap B) + \cdots + Pr(A_K \cap B) \\
&= Pr(A_1)Pr(B \mid A_1) + \cdots + Pr(A_K)Pr(B \mid A_K) \\
&= \sum_{i=1}^{K} Pr(A_i)Pr(B \mid A_i)
\end{aligned}
$$

Bayes Theorem

Statement

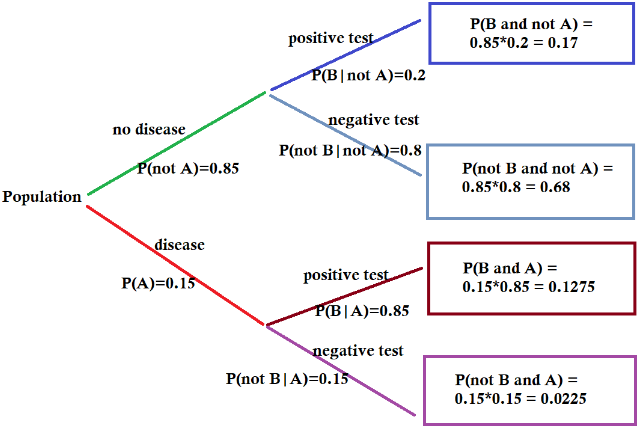
Given a set of events $\{A_1, \ldots, A_i, \ldots, A_j, \ldots, A_K\}$, such that $A_i \cap A_j = \emptyset$, for all $i, j = 1, \ldots, K$, and such that $\cup_{i=1}^{K} A_i = \Omega$. Further, given another event $B \subset \Omega$, we have that, for $i = 1, \ldots, K$:

$$Pr(A_i \mid B) = \frac{Pr(A_i) Pr(B \mid A_i)}{\sum_{i=1}^{K} Pr(A_i) Pr(B \mid A_i)}$$

Bayes Theorem

Example

What is the probability that, if a patient tests positive, then the patient actually has the disease?



Bayes Theorem

Example

- $P(A) = 0.15$ is the probability of having the disease
- $P(\bar{A}) = 0.85$ is the probability of not having the disease
- $P(\bar{B} \mid A) = 0.15$ is the probability of a false negative
- $P(B \mid \bar{A}) = 0.2$ is the probability of a false positive

$$
\begin{aligned}
P(A \mid B) &= \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})} \\
&= \frac{0.15 \times 0.85}{0.15 \times 0.85 + 0.85 \times 0.2} \\
&= \frac{0.1275}{0.1275 + 0.17} = \frac{0.1275}{0.2975} = 0.4286
\end{aligned}
$$

The probability of testing positive given that a patient actually has the disease is not large, but only 0.4286.

Bayes Theorem

Give it a go!

Keepers in a panda sanctuary noticed that 45% of their pandas enjoy climbing trees after having snacked on bamboo, while the remaining ones prefer taking a nap straight away. Half an hour after the meal, 90% of the climbers will opt for a good nap as well. They'll be joining 80% of the pandas that went for a nap in the first place, and are still asleep 30 minutes after having finished eating their bamboo. A panda is selected at random 30 minutes and a second after having eaten bamboo, and it's found sleeping. What is the probability that this panda climbed a tree after the meal, before falling asleep?

[0.4793]

The Monty Hall Problem

Consider a Game Show hosted by Monty Hall. He shows you three doors: A, B, and C. Behind one door, is a fabulous car, behind the others, two goats!

1. You choose one door
2. Monty then opens a door you did not choose to reveal a goat. Monty knows where the car is and will never reveal it. Also, if you picked the car, he is equally likely to pick either goat
3. Monty asks you – would you like to stick with the door you chose, or switch to the other remaining door?

The Monty Hall Problem

How do we solve it?

Context: There are two options: switch or not switch. The question is, should we switch?

Solution

- Assume we choose door A, and that Monty opens door B.
- Let H be the hypothesis that door A contains the car.
- Let E be the evidence that Monty has revealed door B with a goat behind it.
- What is the probability that the car is in A? That is:

$$Pr(\text{H} \mid \text{B open} \cap \text{C closed}) = Pr(\text{H} \mid \text{E})$$

- We can compute this probability by using Bayes Theorem

The Monty Hall Problem

How do we solve it?

$$Pr(\text{H} \mid \text{E}) = \frac{Pr(H)Pr(E \mid H)}{Pr(E)}$$
$$= \frac{Pr(H)Pr(E \mid H)}{Pr(H)Pr(E \mid H) + Pr(\bar{H})Pr(E \mid \bar{H})}$$

where, P(H) = Probability of a car behind door A
$P(\bar{H})$ = Probability that a car is not behind door A
$P(E|H)$ = Probability that Monty shows a goat behind the remaining doors, given the car is behind door A
$P(E|\bar{H})$ = Probability that Monty shows a goat behind the remaining doors, given the car is not behind door A

The Monty Hall Problem

How do we solve it?

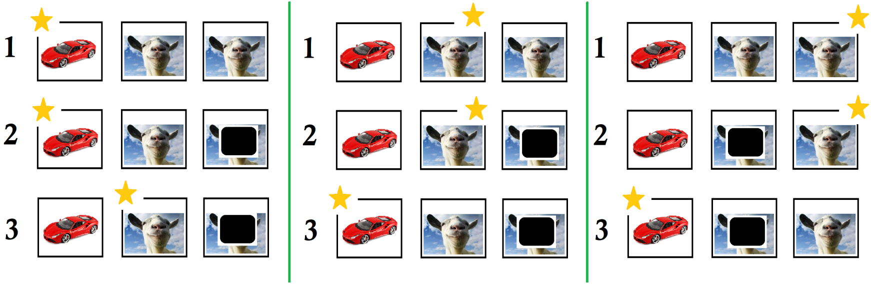
- $\mathrm{P(H)} = \frac{1}{3}$
- $P(\bar{H}) = = \frac{2}{3}$
- $P(E|H) = 1$
- $P(E|\bar{H}) = 1$

Therefore, the probability the car is behind door A, given it is not behind door B is:

$$Pr(\mathrm{H} \mid \mathrm{E}) = \frac{\frac{1}{3} \times 1}{\frac{1}{3} \times 1 + \frac{2}{3} \times 1} = \frac{1}{3}$$

The probability it is behind the door that is not revealed is $\frac{2}{3}$. Therefore, switching is twice as likely to get you the car as staying!

The Monty Hall Problem



shorturl.at/hioHT

Probability - different philosophies

Frequentist paradigm

The concept of probability is linked to the idea of experiments being replicable. Hence, probabilities are computed as relative frequencies: $Pr(A) \approx f_A = \frac{n_A}{n}$. That is, the probability of event A is estimated as the number of times it happened (n_A), over a total of n experiments. Note that here the assumption is that all experiments can be replicated in the exact same conditions. The frequentist approach is then based on the **repeated sampling** principle.

Example

	Blue	Green	Total
Square	350	350	700
Circle	250	50	300
Total	600	400	1000

$$f_{\text{Blue}} = \frac{600}{1000} = 0.6$$

$$f_{Circle,Green} = \frac{50}{1000} = 0.05$$

Probability - different philosophies

Subjective paradigm

The probability of an event A measures the "level of confidence" (not to be confused with confidence intervals!) a coherent individual attributes to the happening of said event, given all information available to them.

One way to interpret this paradigm is, if a bet was to be placed on A, then $Pr(A)$ would be the unitary price one would consider fair to pay if A happened.

The **Bayesian** inferential paradigm is based on the subjective probability paradigm. Very briefly, this approach makes use of "prior" probabilities, alongside with observed data, to try to quantify, summarize, and utilize the knowledge a priori one has on the type of variables and population that are being analysed in the observed data.

Exercises

Exercises

- Consider the toss of a balanced coin.
  - What are the elementary events here?
  - Consider tossing the coin 3 times. What is the probability of getting at least once head?
- If $Pr(A) = 0.8$, what is $Pr(\bar{A})$?
- If $Pr(A) = 0.4$ and $Pr(B) = 0.5$, and A and B are incompatible, what is $Pr(A \cap B)$?
- If $Pr(B) = 1$ and $Pr(A) = 0$, what is $Pr(A \cup B)$? What is $Pr(A \cap B)$?

Exercises

- A box contains 7 cinnamon biscuits, 10 caramel ones, and 20 chocolate ones.
  - What is the probability of getting a chocolate biscuit?
  - What is the probability of getting either a chocolate or a caramel biscuit?
- Consider a bag containing 18 marbles, 6 red and 12 white. We extract 2 marbles **without replacement**, and we'd like to compute the probability that the first marble is red (event A), while the second marble is also red (event B).
- What is the probability of getting a number larger than 10 when rolling two dices?

Exercises

- Consider the data in the following table:

	Small	**Medium**	**Large**	**Total**
Red	150	10	400	560
Green	200	120	30	350
Total	350	130	430	910

  - Compute $Pr(\text{Large})$
  - Compute $Pr(\text{Large} \cap \text{Green})$
  - Compute $Pr(\text{Red} \cup \text{Medium})$
  - Compute $Pr(\text{Red}|\text{Small})$