Probability

Independent Events | Conditional Probability

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Independent Events

- In probability, we say two events are **independent** if **knowing one event occurred doesn't change the probability of the other event**
  - The probability that a fair coin shows "heads" after being flipped is 1/2. What if we knew the day was Tuesday?
  - What about **gender and handedness** (left-handed vs. right-handed)?. When we look at probabilities though, we see that about 10% of all people are left-handed, but about 12 % of males are left-handed.
  - So these events are **not independent**, since **knowing a random person is a male increases the probability that they are left-handed**

Fair Coin

H T

$P(H) = 0.5$

10 %

male ⇒ 12 %

A B

$P(A)$

happened

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

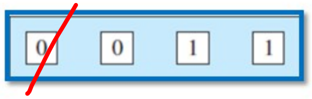
Independent Events

Independence

- The **items in a sample are said to be independent** if knowing the values of some of them does not help to predict the values of the others.
- With a **finite**, **tangible population**, the items in a simple random sample **are not strictly independent**, because as each **item is drawn**, the **population changes**.
- This change can be **substantial** when the **population is small**.
- However, when the **population is very large**, this **change is negligible,** and the **items can be treated as if they were independent**.

Independence

- To illustrate this idea, imagine that we **draw a simple random sample of 2 items** from the population
- **For the first draw**, the numbers **0 and 1 are equally likely**.
- But the **value of the second item** is clearly **influenced** by the first; if the first is 0, the second is more likely to be 1, and vice versa. Thus the **sampled items are dependent**



$p(1) = 2/3$

Independence

One million $\boxed{0}$'s One million $\boxed{1}$'s

- With the **large population**, the sample items are for all practical purposes **independent**.
- It is **reasonable to wonder how large a population must be** in order that the **items** in a simple random sample may be treated as **independent**.

Independence

- Interestingly, it is **possible to make a population behave** as though it were **infinitely large**, by **replacing** each item after it is sampled.
- This method is called **sampling with replacement**.

Independent Events

- Sometimes the knowledge that one event has occurred does not change the probability that another event occurs. In this case the **conditional and unconditional** probabilities are the **same**, and the events are said to be **independent**

$$P(A) = P(A|B) \Rightarrow \text{independent events}$$

Example

Length	Diameter Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

$\Rightarrow 40$

- If an aluminum rod is sampled from the sample space presented in Table above, find P(too long) and P(too long | too thin). Are these probabilities different?

Independent event

$$P(\text{too long}) = \frac{40}{1000} = \boxed{0.04}$$

$$P(\text{too long} \mid \text{too thin}) = \frac{P(\overset{\text{long}}{\text{—}} \cap \overset{\text{thin}}{\text{—}})}{P(\text{too thin})}$$

$$= \frac{2/1000}{50/1000}$$

$$= 2/50 = \boxed{0.04}$$

Independent Events

- Two events A and B are **independent** if the probability of each event remains the same whether or not the other occurs.
- In symbols: If $P(A) \neq 0$ and $P(B) \neq 0$, then A and B are **independent if**

$\boldsymbol{P(B|A) = P(B)}$ or, equivalently, $\boldsymbol{P(A|B) = P(A)}$

If either $P(A) = 0$ or $P(B) = 0$, then **A and B are independent**.

Example - Income and universities

- Researchers surveyed recent graduates of **two different universities** about their **annual incomes**. The following two-way table displays data for the **300 graduates** who responded to the survey

Annual income	University A	University B	TOTAL
Under $20,000	36	24	60
$20,000 to 39,999	109	56	165
$40,000 and over	35	40	75
TOTAL	180	120	300

Suppose we choose a random graduate from this data. **Are the events "income is $40,000 and over" and "attended University B" independent?**

$P(40,000 \text{ and over}) = \frac{75}{300} = 0.25$

$P(40,000 \text{ and over} \mid B) = \frac{40}{120} = 0.333$

$P(40 \text{ and over}) \neq$

$P(40 \text{ and over} \mid B)$

Not Independent

Example - Income and universities

- Researchers surveyed recent graduates of **two different universities** about their **annual incomes**. The following two-way table displays data for the **300 graduates** who responded to the survey

~~Annual income~~	University A	University B	TOTAL
Under $20,000	36	24	60
$20,000 to 39,999	109	56	165
$40,000 and over	35	40	75
TOTAL	180	120	300

Are the events "income under $20,000" and "attended University B" independent?

Prediction // classification

Example

- James is interested in the **weather conditions** and whether the **downtown train** he sometimes takes runs **on times**. For a year, James records whether each day is **Sunny**, **Cloudy**, **Rainy**, or **Snowy**, as well as whether this train **arrives on time or is delayed**. His results are displayed in the table
- For these days, are the events "Delayed" and "Snowy" Independent?

feature label

Weather	**On-Time**	**Delayed**	Total
Sunny	167	3	170
Cloudy	115	5	120
Rainy	40	15	55
Snowy	8	12	20
Total	330	35	365

Not Independent

Feature Selection

Data Science

$$P(\text{delayed}) = \frac{35}{365} = 0.09 \leftarrow 1\%$$

$$P(\text{delayed} \mid \text{snowy}) = \boxed{\frac{P(\text{delayed} \cap \text{snowy})}{P(\text{snowy})}} = \frac{12/365}{20/365} = 0.6$$

$$= 12/20 = 0.6 \quad 60\%$$

References

- Statistics for Engineers and Scientists, Fourth Edition, William Navidi
  Colorado School of Mines
- Khan Academy