Why Statistics in Data Science?

We need grounded means for reasoning about data generated from real world with some degree of randomness.

What will you learn?

- Probability: properties of data generated by a known/assumed randomness model
- Statistics: properties of a randomness model that could have generated given data
- The R programming language
Sample spaces and events

- An **experiment** is a measurement of a random process
- The **outcome** of an experiment takes values in some set $\Omega$, called the **sample space**.

**Examples:**
- Tossing a coin: $\Omega = \{H, T\}$
- Month of birthdays $\Omega = \{Jan, \ldots, Dec\}$
- Population of a city $\Omega = \mathbb{N} = \{0, 1, 2, \ldots, \}$ [Countably infinite sample space]
- Length of a street $\Omega = \mathbb{R}^+ = (0, \infty)$ [Uncountably infinite sample space]
- Tossing a coin twice: $\Omega = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}$
- Testing for Covid-19 (univariate): $\Omega = \{+, -\}$
- Testing for Covid-19 (multivariate): $\Omega = \{f, m\} \times \mathbb{N} \times \{+, -\}$, e.g., $(f, 25, -) \in \Omega$

- An **event** is some subset of $A \subseteq \Omega$ of possible outcomes of an experiment.
  - $L = \{Jan, March, May, July, August, October, December\}$ a long month with 31 days
- We say that an event $A$ **occurs** if the outcome of the experiment belongs to the set $A$.
  - If the outcome is Jan then $L$ occurs

  Look at [seeing-theory.brown.edu](http://seeing-theory.brown.edu)
Probability functions on finite sample space

A **probability function** is a mapping from events to **real numbers** that satisfies certain axioms. *Intuition: how likely is an event to occur.*

**Definition.** A *probability function* $P$ on a finite sample space $\Omega$ assigns to each event $A$ in $\Omega$ a number $P(A)$ in $[0,1]$ such that

(i) $P(\Omega) = 1$, and

(ii) $P(A \cup B) = P(A) + P(B)$ if $A$ and $B$ are disjoint.

The number $P(A)$ is called the probability that $A$ occurs.

- **Fact:** $P(\{a_1, \ldots, a_n\}) = P(\{a_1\}) + \ldots + P(\{a_n\})$  
  [Generalized additivity]

- **Assigning probability to a singleton is enough**

- **Examples:**
  - $P(\{H\}) = P(\{T\}) = \frac{1}{2}$
  - $P(\{\text{Jan}\}) = \frac{31}{365}, P(\{\text{Feb}\}) = \frac{28}{365}, \ldots P(\{\text{Dec}\}) = \frac{31}{365}$
  - $P(L) = \frac{7}{12}$ or $31\cdot\frac{7}{365}$?

- **$P(\{a\})$** often abbreviated as $P(a)$, e.g., $P(\text{Jan})$ instead of $P(\{\text{Jan}\})$
Properties of probability functions

- \( P(A^c) = 1 - P(A) \)
- \( P(\emptyset) = 0 \)
- \( A \subseteq B \Rightarrow P(A) \leq P(B) \) \[Impossible event\] \[Monotonicity\]
- \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \) \[Inclusion-exclusion principle\]
- Example: \( P(A \cup B) = P(A) + P(B \setminus A) \)
- probability that at least one coin toss over two lands head?
  - Tossing a coin twice: \( \Omega = \{H, T\} \times \{H, T\} = \{(H, H),(H, T),(T, H),(T, T)\} \)
  - \( A = \{(H, H),(H, T)\} \) first coin is head
  - \( B = \{(H, H),(T, H)\} \) second coin is head
  - Answer \( P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4} \)
Assigning probability is **NOT** an easy task: a prob. function can be an approximation of reality

- **Frequentist** interpretation: probability measures a “proportion of outcomes”.
  - A fair coin lands on heads 50% of times
  - \[ P(A) = \frac{|A|}{|\Omega|} \]
  - \[ P(\{ \text{at least one H in two coin tosses}\}) = \frac{|\{(H, H), (H, T), (T, H)\}|}{4} = \frac{3}{4} \]

- **Bayesian** (or epistemological) interpretation: probability measures a “degree of belief”.
  - (We believe that) Iliad and Odyssey were written by the same person at 90%
Probability functions on countably infinite sample space

DEFINITION. A probability function on an infinite (or finite) sample space $\Omega$ assigns to each event $A$ in $\Omega$ a number $P(A)$ in $[0, 1]$ such that

(i) $P(\Omega) = 1$, and
(ii) $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$

if $A_1, A_2, A_3, \ldots$ are disjoint events.

(ii) is called **countable additivity**. It is equivalent to $\sigma$-additivity: for $A_1 \subseteq A_2 \subseteq \ldots$

$$P(\lim_{n \to \infty} A_i) = \lim_{n \to \infty} P(A_i)$$

Example

- Experiment: we toss a coin repeatedly until H turns up.
- Outcome: the number of tosses needed.
- $\Omega = \{1, 2, \ldots\} = \mathbb{N}^+$
- Suppose: $P(H) = p$. Then: $P(n) = (1 - p)^{n-1}p$
- Is it a probability function? $P(\Omega) = \ldots$
Conditional probability

- Long months and months with ‘r’
  - $L = \{ \text{Jan, Mar, May, July, Aug, Oct, Dec} \}$ — a long month with 31 days
  - $R = \{ \text{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec} \}$ — a month with ‘r’
  - $P(L) = \frac{7}{12}$  \hspace{1cm} $P(R) = \frac{8}{12}$

- Anna is born in a long month. What is the probability she is born in a month with ‘r’?

  $$P(R|L) = \frac{P(L \cap R)}{P(L)} = \frac{P(\{\text{Jan, Mar, Oct, Dec}\})}{P(L)} = \frac{4/12}{7/12} = \frac{4}{7}$$

- **Intuition:** probability of an event in the restricted sample space $\Omega \cap L$
  - *a-priori* probability $P(R) = \frac{8}{12}$
  - *a-posteriori* probability $P(R|L) = \frac{4}{7} < \frac{8}{12}$
Conditional probability

**Definition.** The *conditional probability* of $A$ given $C$ is given by:

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

provided $P(C) > 0$.

**Properties:**
- $P(A | C) \neq P(C | A)$, in general
- $P(\Omega | C) = 1$
- if $A \cap B = \emptyset$ then $P(A \cup B | C) = P(A | C) + P(B | C)$

**The Multiplication Rule:** For any events $A$ and $C$:

$$P(A \cap C) = P(A | C) \cdot P(C).$$

More generally, the **Chain Rule**:

$$P(A_1 \cap A_2 \cap A_3 \ldots \cap A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot \ldots \cdot P(A_n | \cap_{i=1}^{n-1} A_i)$$
Example: no coincident birthdays

- \( B_n = \{ n \text{ different birthdays} \} \)
- For \( n = 1 \), \( P(B_1) = 1 \)
- For \( n > 1 \),
  \[
P(B_n) = P(B_{n-1}) \cdot P(\{\text{the } n\text{-th person's birthday differs from the other } n-1\}|B_{n-1})
  = P(B_{n-1}) \cdot (1 - \frac{n-1}{365}) = \ldots = \prod_{i=1}^{n-1} \left(1 - \frac{i}{365}\right)
  \]
The law of total probability

**The Law of Total Probability.** Suppose $C_1, C_2, \ldots, C_m$ are disjoint events such that $C_1 \cup C_2 \cup \cdots \cup C_m = \Omega$. The probability of an arbitrary event $A$ can be expressed as:

$$P(A) = P(A \mid C_1)P(C_1) + P(A \mid C_2)P(C_2) + \cdots + P(A \mid C_m)P(C_m).$$

- **Intuition:** case-based reasoning

---

**Fig. 3.2.** The law of total probability (illustration for $m = 5$).
Example: case-based reasoning

Factory 1’s light bulbs work for over 5000 hours in 99% of cases.  
Factory 2’s bulbs work for over 5000 hours in 95% of cases.  
Factory 1 supplies 60% of the total bulbs on the market and Factory 2 supplies 40% of it.  

**Question:** What is the chance that a purchased bulb will work for longer than 5000 hours?

- $A = \{ \text{bulbs working for longer than 5000 hours} \}$
- $C_1 = \{ \text{bulbs made by Factory 1} \}$, hence $C_2 = \{ \text{bulbs made by Factory 2} \}$
- Since $\Omega = C_1 \cup C_2$ and $C_1 \cap C_2 = \emptyset$, by the multiplication rule:

$$P(A) = P(A|C_1) \cdot P(C_1) + P(A|C_2) \cdot P(C_2)$$

**Answer:** $P(A) = 0.99 \cdot 0.6 + 0.95 \cdot 0.4 = 0.974$
Example: The Monty Hall problem

https://www.mathwarehouse.com/monty-hall-simulation-online/
(See also Exercise 2.14 of textbook [T])

Tree-based sequential description of probability function
Assume player choose Door 1

| Car location: | Host opens: | Total probability: | Stay: | Switch: |
|---------------|------------|--------------------|-------|---------|
| Door 1        | 1/2        | 1/6                | Car   | Goat    |
| 1/3           |            |                    |       |         |
| Door 2        | 1/2        | 1/6                | Car   | Goat    |
| 1/3           |            |                    |       |         |
| Door 3        | 1/2        | 1/6                | Goat  | Car     |
| 1/3           |            |                    |       |         |
| Door 3        | 1/2        | 1/6                | Goat  | Car     |
| 1/3           |            |                    |       |         |
Independence of events

**Intuition:** whether one event provides any information about another.

An event $A$ is independent of $B$, if $P(B) = 0$ or $P(A|B) = P(A)$

- For $P(R|L) = \frac{4}{7} \neq \frac{8}{12} = PR(R)$ - knowing Anna was born in a long month change the probability she was born in a month with ’r’!
- Tossing 2 coins:
  - $A_1$ is “H on toss 1” and $A_2$ is “H on toss 2”
  - $P(A_1) = P(A_2) = \frac{1}{2}$
  - $P(A_2|A_1) = P(A_2 \cap A_1)/P(A_1) = \frac{1}{4}/\frac{1}{2} = \frac{1}{2} = P(A_1)$
- Properties:
  - $A$ independent of $B$ iff $P(A \cap B) = P(A) \cdot P(B)$
  - $A$ independent of $B$ iff $B$ independent of $A$ \(\textbf{[Symmetry]}\)
  - $A$ independent of $B$ iff $A^c$ independent of $B$
Physical independence and stochastic independence

Independence

An event $A$ is independent of $B$, if $P(B) = 0$ or $P(A|B) = P(A)$

- Physical independence implies stochastic independence
  - However, physical independence is quite a subtle matter (see the butterfly effect)
- But there are stochastic independent events that are physically dependent
  - Suppose a fair die is rolled twice.
  - $A$ = “a three is obtained on the second roll”
  - $B$ = “the sum of the two numbers obtained is less than or equal to 4”
  - Exercise at home. Prove that $P(A|B) = P(A)$
Conditional independence of events

**Intuition:** whether one event provides any information about another given a third event occurred. Technically, consider $P(\cdot|C)$ in independence.

### Conditional independence

An event $A$ is conditionally independent of $B$ given $C$ such that $P(C) > 0$, if $P(B|C) = 0$ or

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

- **Properties:**
  - $A$ conditionally independent of $B$ iff $P(A \cap B|C) = P(A|C) \cdot P(B|C)$
  - $A$ conditionally independent of $B$ iff $B$ conditionally independent of $A$ \(\text{[Symmetry]}\)

- **Exercise at home.** Prove or disprove:
  - If $A$ is independent of $B$ then $A$ is conditionally independent of $B$ given $C$
Independence of two or more events

Events $A_1$, $A_2$, $\ldots$, $A_m$ are called independent if

$$P(A_1 \cap A_2 \cap \cdots \cap A_m) = P(A_1)P(A_2)\cdots P(A_m)$$

and this statement also holds when any number of the events $A_1$, $\ldots$, $A_m$ are replaced by their complements throughout the formula.

Alternative definition

Events $A_1$, $A_2$, $\ldots$, $A_m$ are called independent if for every $J \subseteq \{1, \ldots, m\}$:

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$$

Exercise at home: show the two definitions are equivalent
Independence of two or more events

Alternative definition

Events $A_1, A_2, \ldots, A_m$ are called independent if for every $J \subseteq \{1, \ldots, m\}$:

$$P\left( \bigcap_{i \in J} A_i \right) = \prod_{i \in J} P(A_i)$$

• It is **stronger** than pairwise independence

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \text{ for } i \neq j \in \{1, \ldots, m\}$$

• Example: what is the probability of at least one head in the first 10 tosses of a coin?

$A_i = \{\text{head in } i\text{-th toss}\}$

$$P\left( \bigcup_{i=1}^{10} A_i \right) = 1 - P\left( \bigcap_{i=1}^{10} A_i^c \right) = 1 - \prod_{i=1}^{10} P(A_i^c) = 1 - \prod_{i=1}^{10} (1 - P(A_i))$$