University of BRISTOL

Introduction to probability theory

Henry W J Reeve

henry.reeve@bristol.ac.uk

Statistical Computing & Empirical Methods (EMATM0061)

MSc in Data Science, Teaching block 1, 2021.

What will we cover today?

- We will consider data sets as random samples from a population of interest.
- We introduced the formal concept of probability;
- We will derive several important consequences of the rules of probability;
- We will briefly discuss the different interpretations of probability.

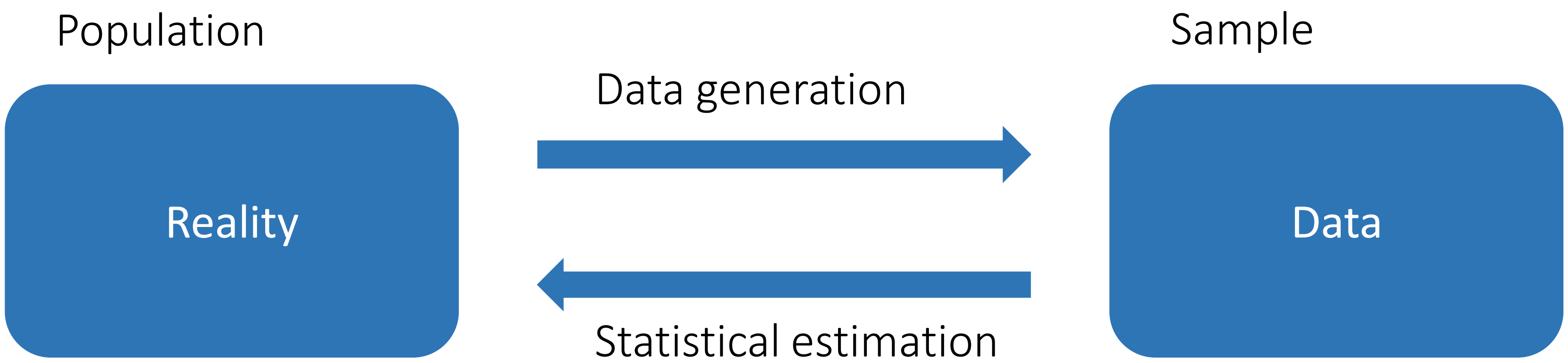
Samples and populations

We attempt to understand populations of penguins by looking at random samples.

Population

Sample

Statistical estimation and probability



To model the stochastic data generation process we will require some probability theory!

Random experiments and sample spaces

A **random experiment** is a procedure (real or imagined) which:

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

An **event** is a set (i.e. a collection) of possible outcomes.

A **sample space** is the set of all possible outcomes of interest for a random experiment.

What is probability?

We often make statements about the **probability**, **likelihood** or **chance** of different events.

"If you like listening to Eric Clapton then you're probably also a fan of Jimi Hendrix".

"There is a good chance that the level of inflation will fall due to the rise in interest rates."

"Given how cloudy it is, there's a high likelihood it will rain."

"Bristol City Football Club probably won't win the FA Cup this year."

We need **probability theory** to make such statements precise so we can reason about them quantitatively.

The laws of probability

The probability of an event is a numerical value used to quantify how likely it is to occur.

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$, where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

The laws of probability

The probability of an event is a numerical value used to quantify how likely it is to occur.

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$, where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies rules 1,2,3:

The laws of probability

The probability of an event is a numerical value used to quantify how likely it is to occur.

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$, where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies rules 1,2,3:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

The laws of probability

The probability of an event is a numerical value used to quantify how likely it is to occur.

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$,
where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies rules 1,2,3:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

The laws of probability

The probability of an event is a numerical value used to quantify how likely it is to occur.

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$, where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies rules 1,2,3:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every countably infinite sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right).$$

The laws of probability

Given a sample space Ω along with a well-behaved collection of events $\mathcal{E}$,a **probability** $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$, and satisfies rules 1,2,3:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every countably infinite sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}\left(A_i\right).$$

These rules are known as the Kolmogorov axioms after the great Andrey Kolmogorov who formalized them in 1933.

The laws of probability – Example 1

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots$, we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

<u>Example 1</u>

Consider the rolls of a fair dice.

The sample space is $\Omega = \{1,2,3,4,5,6\}$.

The set of events is $\mathcal{E} = \{A \subseteq \Omega\}$, the collection of all subsets.

For any $A \in \mathcal{E}$ we have $\mathbb{P}(A) = \frac{|A|}{6}$.

The laws of probability – Example 2

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Example 2

A customer in dealer ship either buys a car (1) or doesn't buy a car (0).

The sample space is $\Omega = \{0,1\}$.

The set of events is $\mathcal{E} = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$, the collection of all subsets.

Let $p \in (0,1)$ denote the probability that a purchase is made then,

$$\mathbb{P}(A) := \begin{cases} 0 & \text{if } A = \emptyset \\ 1-p & \text{if } A = \{0\} \\ p & \text{if } A = \{1\} \\ 1 & \text{if } A = \{0,1\}. \end{cases}$$

The empty set has zero probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots$, we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

The empty set has zero probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

Proof: To see this we consider the infinite sequence $A_1, A_2, \ldots,$ where $A_i = \emptyset$ for all $i \in \mathbb{N}$.

Observe that for any $i \neq j \in \mathbb{N}$ we have $A_i \cap A_j = \emptyset \cap \emptyset = \emptyset$,so the sequence $A_1, A_2, \ldots,$ is pairwise disjoint.

The empty set has zero probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0.$

Proof: To see this we consider the infinite sequence $A_1, A_2, \dots,$ where $A_i = \emptyset$ for all $i \in \mathbb{N}$.

Observe that for any $i \neq j \in \mathbb{N}$ we have $A_i \cap A_j = \emptyset \cap \emptyset = \emptyset$,so the sequence $A_1, A_2, \dots,$ is pairwise disjoint.

By the third rule we have $\mathbb{P}(\emptyset) = \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$.

The empty set has zero probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0.$

Proof: To see this we consider the infinite sequence $A_1, A_2, \dots,$ where $A_i = \emptyset$ for all $i \in \mathbb{N}$.

Observe that for any $i \neq j \in \mathbb{N}$ we have $A_i \cap A_j = \emptyset \cap \emptyset = \emptyset$,so the sequence $A_1, A_2, \dots,$ is pairwise disjoint.

By the third rule we have $\mathbb{P}(\emptyset) = \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$.

This is only possible if $\mathbb{P}(\emptyset) = 0$.

The monotonicity property of probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

The monotonicity property of probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proof: Consider the infinite sequence $S_1, S_2, \dots,$ where $S_1 = A$, $S_2 = B \setminus A$,and $S_i = \emptyset$ for all $i \geq 3$.

The monotonicity property of probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proof: Consider the infinite sequence $S_1, S_2, \dots,$ where $S_1 = A$, $S_2 = B \backslash A$,and $S_i = \emptyset$ for all $i \geq 3$.

Observe that for any $i \neq j \in \mathbb{N}$ we have $S_i \cap S_j = \emptyset$,so the sequence $S_1, S_2, \dots,$ is pairwise disjoint.

The monotonicity property of probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proof: Consider the infinite sequence $S_1, S_2, \ldots,$ where $S_1 = A$, $S_2 = B \backslash A$,and $S_i = \emptyset$ for all $i \geq 3$.

Observe that for any $i \neq j \in \mathbb{N}$ we have $S_i \cap S_j = \emptyset$,so the sequence $S_1, S_2, \ldots,$ is pairwise disjoint.

Observe that since $A \subseteq B$ we have $B = \mathrm{A} \cup (B \backslash A)$.

The monotonicity property of probability

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proof: Consider the infinite sequence $S_1, S_2, \ldots,$ where $S_1 = A$, $S_2 = B \backslash A$,and $S_i = \emptyset$ for all $i \geq 3$.

Observe that for any $i \neq j \in \mathbb{N}$ we have $S_i \cap S_j = \emptyset$,so the sequence $S_1, S_2, \ldots,$ is pairwise disjoint.

Observe that since $A \subseteq B$ we have $B = \mathrm{A} \cup (B \backslash A)$. Hence, by the third and then the first rule we have

$$\mathbb{P}(B) = \mathbb{P}(A \cup (B \backslash A)) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} S_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(S_i) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) + \sum_{i=1}^{\infty} \mathbb{P}(\emptyset) \geq \mathbb{P}(A).$$

Probabilities are between zero and one

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 3: If $A \in \mathcal{E}$ is any event, then $0 \leq \mathbb{P}(A) \leq 1$.

Probabilities are between zero and one

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots$, we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 3: If $A \in \mathcal{E}$ is any event, then $0 \leq \mathbb{P}(A) \leq 1$.

Proof: For any event $A \in \mathcal{E}$ we have $A \subseteq \Omega$.

Probabilities are between zero and one

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 3: If $A \in \mathcal{E}$ is any event, then $0 \leq \mathbb{P}(A) \leq 1$.

Proof: For any event $A \in \mathcal{E}$ we have $A \subseteq \Omega$.

Thus, by Consequence 2 and Rule 2 we have $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$.

Probabilities are between zero and one

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 3: If $A \in \mathcal{E}$ is any event, then $0 \leq \mathbb{P}(A) \leq 1$.

Proof: For any event $A \in \mathcal{E}$ we have $A \subseteq \Omega$.

Thus, by Consequence 2 and Rule 2 we have $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$.

We also have $\mathbb{P}(A) \geq 0$ by rule 1.

The union bound

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 4: Given any sequence of sets $S_1, S_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

The union bound

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 4: Given any sequence of sets $S_1, S_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Proof: Define a sequence of sets $A_1, A_2, \ldots,$ by $A_1 := S_1$, $A_2 := S_2 \backslash S_1$, $A_3 := S_3 \backslash (S_1 \cup S_2)$ and so on with

$A_i := S_i \backslash \left(\bigcup_{j<i} S_j \right) = S_i \backslash (S_1 \cup S_2 \cup \cdots \cup S_{i-1})$.

The union bound

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 4: Given any sequence of sets $S_1, S_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Proof: Define a sequence of sets $A_1, A_2, \dots,$ by $A_1 := S_1$, $A_2 := S_2 \backslash S_1$, $A_3 := S_3 \backslash (S_1 \cup S_2)$ and so on with

$A_i := S_i \backslash \left(\bigcup_{j<i} S_j\right) = S_i \backslash (S_1 \cup S_2 \cup \dots \cup S_{i-1})$.

It follows that $A_{i_1} \cap A_{i_0} \subseteq \left\{S_{i_1} \backslash \left(\bigcup_{j<i_1} S_j\right)\right\} \cap S_{i_0} = \emptyset$ whenever $i_0 < i_1$.

The union bound

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 4: Given any sequence of sets $S_1, S_2, \dots,$ we have $\mathbb{P}(\cup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Proof: Define a sequence of sets $A_1, A_2, \dots,$ by $A_1 := S_1$,$A_2 := S_2 \backslash S_1$,$A_3 := S_3 \backslash (S_1 \cup S_2)$ and so on with

$A_i := S_i \backslash \left(\cup_{j<i} S_j\right) = S_i \backslash (S_1 \cup S_2 \cup \cdots \cup S_{i-1})$.

It follows that $A_{i_1} \cap A_{i_0} \subseteq \left\{S_{i_1} \backslash \left(\cup_{j<i_1} S_j\right)\right\} \cap S_{i_0} = \emptyset$ whenever $i_0 < i_1$.

Hence, $A_1, A_2, \dots,$ are pairwise disjoint.

The union bound

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 4: Given any sequence of sets $S_1, S_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Proof: Define a sequence of sets $A_1, A_2, \ldots,$ by $A_1 := S_1$, $A_2 := S_2 \backslash S_1$, $A_3 := S_3 \backslash (S_1 \cup S_2)$ and so on with

$A_i := S_i \backslash \left(\bigcup_{j<i} S_j\right) = S_i \backslash (S_1 \cup S_2 \cup \cdots \cup S_{i-1})$.

It follows that $A_{i_1} \cap A_{i_0} \subseteq \left\{S_{i_1} \backslash \left(\bigcup_{j<i_1} S_j\right)\right\} \cap S_{i_0} = \emptyset$ whenever $i_0 < i_1$.

Hence, $A_1, A_2, \ldots,$ are pairwise disjoint.

Note also that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \left\{S_i \backslash \left(\bigcup_{j<i} S_j\right)\right\} = \bigcup_{i=1}^{\infty} S_i$.

The union bound

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots$, we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 4: Given any sequence of sets $S_1, S_2, \dots$, we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Proof: Define a sequence of sets $A_1, A_2, \dots$, by $A_1 := S_1$, $A_2 := S_2 \setminus S_1$, $A_3 := S_3 \setminus (S_1 \cup S_2)$ and so on with

$A_i := S_i \setminus \left(\bigcup_{j<i} S_j\right) = S_i \setminus (S_1 \cup S_2 \cup \dots \cup S_{i-1})$.

It follows that $A_{i_1} \cap A_{i_0} \subseteq \left\{S_{i_1} \setminus \left(\bigcup_{j<i_1} S_j\right)\right\} \cap S_{i_0} = \emptyset$ whenever $i_0 < i_1$.

Hence, $A_1, A_2, \dots$, are pairwise disjoint.

Note also that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \left\{S_i \setminus \left(\bigcup_{j<i} S_j\right)\right\} = \bigcup_{i=1}^{\infty} S_i$.

Hence, by Rule 3, followed by Consequence 2, $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) = \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

The laws of probability and their consequences

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequence 1: $\mathbb{P}(\emptyset) = 0$.

Consequence 2: If $A, B \in \mathcal{E}$ are events with $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

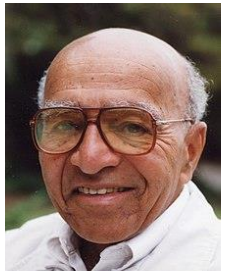
Consequence 3: If $A \in \mathcal{E}$ is any event, then $0 \leq \mathbb{P}(A) \leq 1$.

Consequence 4: Given any sequence of sets $S_1, S_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Interpretations of probability

Thomas Bayes

David Blackwell



Peter Green

There are many different popular interpretations of probability.

<u>Epistemic interpretations:</u>

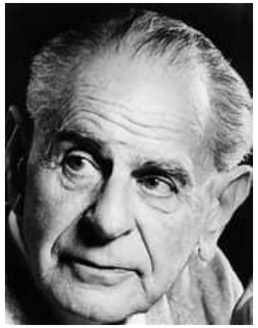
The probability of an event corresponds to a reasonable **degree of belief** for either an individual or a community.

Especially reasonable for one off events e.g. “The probability that Liverpool FC will win the league this year”.

Motivates a **Bayesian** approach to Statistics and Machine Learning.

Interpretations of probability

Jerzy Neyman



Karl Popper

Deborah Mayo

There are many different popular interpretations of probability.

Objective interpretations:

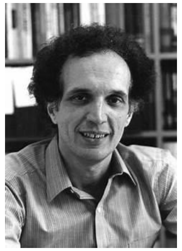
The probability of an event is an objective feature of reality independent of our beliefs.

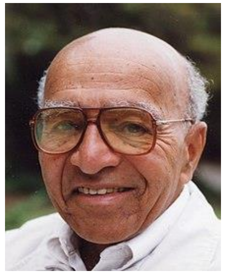
Frequentists view probability as the long run average frequency in the context of a very large number of (possibly hypothetical) replications of an experiment.

The **propensity** view is that probabilities are dispositions to yield certain types of frequency behavior.

Objective interpretations are prevalent in the natural sciences e.g. “The probability an isotope decays in an hour”.

Interpretations of probability

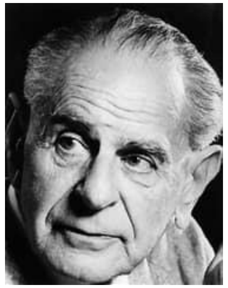




There are many different popular interpretations of probability.


<u>Epistemic interpretations:</u> The probability of an event corresponds to a reasonable degree of belief for either an individual or a community.


<u>Objective interpretations:</u> The probability of an event is an objective feature of reality independent of our beliefs.




Kolmogorov’s laws of probability apply to both.

Both views are valuable in certain contexts and sometimes complementary.

Sigma algebras (an optional extra)

To make the definition of probability rigorous we require the concept of a σ-algebra.

A σ-algebra is a collection $\mathcal{E}$ consisting of subsets $A \subseteq \mathcal{E}$ satisfying:

1. The set $\Omega \in \mathcal{E}$ is contained in $\mathcal{E}$;
2. If $A \in \mathcal{E}$ then $\Omega \backslash A \in \mathcal{E}$;
3. If there is a countable sequence $A_1, A_2, \ldots$ with each $A_i \in \mathcal{E}$,then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$.

Given a sample space Ω along with a σ-algebra of events $\mathcal{E}$,a **probability** $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ such that:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Rule 2: The sample space Ω has probability $\mathbb{P}(\Omega) = 1$.

Rule 3: For every countable sequence of pairwise disjoint events $A_1, A_2, \ldots$, we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

What have we covered?

- We introduced the formal concept of probability as governed by Kolmogorov’s axioms.

- We derived several important consequences of these rules:
  - The empty set has zero probability;
  - Probability is monotonic;
  - The probability of an event is always between zero and one;
  - The union bound.

- We briefly discussed the different interpretations of probability.

University of BRISTOL

Thanks for listening!

Henry W J Reeve

henry.reeve@bristol.ac.uk

Include EMATM0061 in the subject of your email.