

Origins of probability theory

Sources/motivation:

- games of chance
- physics of gases (statistical physics)
- statistical data analysis

Overarching goal:

Find regularity inside unpredictable or chaotic phenomena

A **probability space** is a triplet $(\Omega, \mathcal{F}, P)$ where

- Ω is the set of possible outcomes called **sample space**
- $\mathcal{F}$ is a set of **events** (technically, subsets of Ω)
- P is a function $\mathcal{F} \rightarrow [0, 1]$ with $P(A) :=$ **probability of A**

$P(A)$ quantifies the chance, likelihood, frequency, odds ... that A **occurs**

- **coin toss:** $\Omega := \{0, 1\}$, $\mathcal{F} := \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ and

$$P(\emptyset) := 0, P(\{0\}) := \frac{1}{2}, P(\{1\}) := \frac{1}{2}, P(\{0, 1\}) := 1$$

- **n coin tosses:** $\Omega = \{0, 1\}^n$, $\mathcal{F} := \mathcal{P}(\Omega)$ and

$$P(A) := \frac{|A|}{2^n}$$

- **rolling a die:** $\Omega := \{1, \dots, 6\}$, $\mathcal{F} := \mathcal{P}(\Omega)$ and

$$P(A) := \frac{|A|}{6}$$

- **rolling a die n -times:** $\Omega := \{1, \dots, 6\}^n$, $\mathcal{F} := \mathcal{P}(\Omega)$ and

$$P(A) := \frac{|A|}{6^n}$$

Note that in all cases:

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

This permits more examples:

- **n dice rolls but observe only 1's:** $\Omega := \{0, 1\}^n$, $\mathcal{F} := \mathcal{P}(\Omega)$ and

$$P(\{\omega\}) := \frac{5^{n-k(\omega)}}{6^n} \quad \text{where} \quad k(\omega) := \sum_{i=1}^n \omega_i$$

- **Bernoulli(p)/unfair coin tosses:** $\Omega := \{0, 1\}^n$, $\mathcal{F} := \mathcal{P}(\Omega)$ and

$$P(\{\omega\}) := \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i}$$

where $p \in [0, 1]$

Assumptions on $\mathcal{F}$:

- $\emptyset, \Omega \in \mathcal{F}$
- $\forall A, B \in \mathcal{F}: A \cap B, A \cup B \in \mathcal{F}$
- $\forall A \in \mathcal{F}: A^c \in \mathcal{F}$

Any $\mathcal{F}$ with these properties is called an **algebra** of subsets of Ω .

Assumptions on P :

- $P(\emptyset) = 0 \wedge P(\Omega) = 1$
- $\forall A, B \in \mathcal{F}: A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Any P with these properties is a **finitely-additive set function** on $\mathcal{F}$

Lemma

Let P be a finitely additive set function on an algebra $\mathcal{F}$. Then $\forall A, B \in \mathcal{F}$:

- $P(A^c) = 1 - P(A)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and so
- $P(A \cup B) \leq P(A) + P(B)$

Proof: First line: $A \cup A^c = \Omega$ with $A \cap A^c = \emptyset$. So $P(A) + P(A^c) = P(\Omega) = 1$. Next, $A \subseteq B \Rightarrow B = A \cup (B \setminus A)$. Observe also that $B \setminus A = B \cap A^c \in \mathcal{F}$ and $A \cap (B \setminus A) = \emptyset$. So

$$P(B) = P(A) + P(B \setminus A) \geq P(A)$$

Finally, $A = (A \setminus B) \cup (A \cap B)$ and $B = (B \setminus A) \cup (A \cap B)$ and also $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. These are disjoint unions and so

$$\begin{aligned} P(A) + P(B) &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) + P(A \cap B) \\ &= P(A \cup B) + P(A \cap B) \quad \square \end{aligned}$$

Union bound: $\forall n \geq 1 \forall A_1, \dots, A_n \in \mathcal{F}$:

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$$

Inclusion-Exclusion formula: $\forall n \geq 1 \forall A_1, \dots, A_n \in \mathcal{F}$:

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq j_1 < \dots < j_k \leq n} P\left(\bigcap_{i=1}^k A_{j_i}\right)$$

Related to this are **Bonferroni inequalities**

Write

$$A_n \downarrow A := \left(\forall n \geq 1: A_{n+1} \subseteq A_n \wedge \bigcap_{n \geq 1} A_n = A \right)$$

It then seems reasonable to assume:

Downward continuity of P : For all events $\{A_n\}_{n \geq 1}: A_n \downarrow \emptyset \Rightarrow P(A_n) \rightarrow 0$

Note that for $A_n \downarrow A$ this gives:

$$P(A_n) = P(A_n \setminus A) + P(A) \rightarrow P(A)$$

because $A_n \setminus A \downarrow \emptyset$, provided $A \in \mathcal{F}$. To avoid the proviso, we need:

A continuity assumption on $\mathcal{F}$: For all events $\{A_n\}_{n \geq 1}$ and all $A \subseteq \Omega$,

$$A_n \downarrow A \Rightarrow A \in \mathcal{F}$$

Lemma

Under the above assumptions:

- $\mathcal{F}$ is (an algebra that is) closed under countable unions and countable intersections, and so is a **σ -algebra** on Ω ,
- P is a countably additive set function with total mass 1 and is thus a **probability measure** on $(\Omega, \mathcal{F})$.

$\mathcal{F}$ is a σ -algebra: Let $\{A_k\}_{k \geq 1}$ be events in $\mathcal{F}$. Then

$$\bigcap_{k=1}^n A_k \downarrow \bigcap_{k \geq 1} A_k$$

and since $\bigcap_{k=1}^n A_k \in \mathcal{F}$ for all $n \geq 1$, we conclude $\bigcap_{k \geq 1} A_k \in \mathcal{F}$. De Morgan then gives also $\bigcup_{k \geq 1} A_k \in \mathcal{F}$.

P is countably additive: Let $\{A_k\}_{k \geq 1}$ be disjoint events; i.e., $\forall k, \ell \geq 1$:

$k \neq \ell \Rightarrow A_k \cap A_\ell = \emptyset$. Set $B_n := \bigcup_{k > n} A_k$. Then $\bigcup_{n \geq 1} A_n = B_n \cup \bigcup_{k=1}^n A_k$ and since this is a finite disjoint union,

$$P\left(\bigcup_{n \geq 1} A_n\right) = P(B_n) + \sum_{k=1}^n P(A_k)$$

Now $B_n \downarrow \emptyset$ thanks to disjointness which gives $P(B_n) \rightarrow 0$ while $\sum_{k=1}^n P(A_k) \rightarrow \sum_{k=1}^{\infty} P(A_k)$ by convergence of sums. Hence we get

$$P\left(\bigcup_{n \geq 1} A_n\right) = \sum_{k=1}^{\infty} P(A_k)$$

as desired. □

A **probability space** is thus a triple $(\Omega, \mathcal{F}, P)$ where $\mathcal{F}$ and P obey:

Definition

$\mathcal{F}$ is a σ -**algebra** on Ω if $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ and

- $\emptyset, \Omega \in \mathcal{F}$
- $\forall A \in \mathcal{F}: A^c \in \mathcal{F}$
- $\forall \{A_n\}_{n \geq 1} \in \mathcal{F}^{\mathbb{N}}: \bigcup_{n \geq 1} A_n \in \mathcal{F}$

Definition

Given a σ -algebra $\mathcal{F}$ on Ω , a map $P: \mathcal{F} \rightarrow \mathbb{R}_+$ is a **probability measure** on $(\Omega, \mathcal{F})$ if $\forall \{A_n\}_{n \geq 1} \in \mathcal{F}^{\mathbb{N}}$:

$$\left(\forall j, k \geq 1: j \neq k \Rightarrow A_k \cap A_j = \emptyset \right) \Rightarrow P\left(\bigcup_{n \geq 1} A_n\right) = \sum_{k=1}^{\infty} P(A_k)$$

and

$$P(\Omega) = 1$$

All probabilistic considerations will be based on this model.

This structure does not come for free. Consider this example:

- **∞ -many coin tosses:** $\Omega := \{0, 1\}^{\mathbb{N}}$, $\mathcal{F}$ is a σ -algebra on Ω containing all the sets of the form

$$\left\{ \omega \in \Omega : \forall i = 1, \dots, n : \omega_i = \eta_i \right\}$$

for $n \geq 1$ and $\eta_1, \dots, \eta_n \in \{0, 1\}$, and P assigns this set measure 2^{-n} ?

Problem: Do such $\mathcal{F}$ and P exist?

This goes under the name **measure extension** problem.