# Mathematical background

Christos Dimitrakakis

June 5, 2024

# Outline

# Logic

## Statements

- A statement $A$ may be true or false

## Unary operators

- negation: $\neg A$ is true if $A$ is false (and vice-versa).

## Binary operators

- or: $A \vee B$ ($A$ or $B$) is true if either $A$ or $B$ are true.
- and: $A \wedge B$ is true if both $A$ and $B$ are true.
- implies: $A \Rightarrow B$: is false if $A$ is true and $B$ is false.
- iff: $A \Leftrightarrow B$: is true if $A, B$ have equal truth values.

## Operator precedence

$\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$

# Set theory

- First, consider some universal set $\Omega$.
- A set $A$ is a collection of points $x$ in $\Omega$.
- $\{x \in \Omega : f(x)\}$: the points in $\Omega$ for which $f(x)$ is true.
- Empty set: $\emptyset$. No point ever belong in the empty set.

## Unary operators

- Negation: $\neg A = \{x \in \Omega : x \notin A\} = \Omega \setminus A$.

## Binary operators

- Union: $A \cup B = \{x \in \Omega : x \in A \vee x \in B\}$ - (c.f. $A \vee B$)
- Intersection: $A \cap B = \{x \in \Omega : x \in A \wedge x \in B\}$ - (c.f. $A \wedge B$)
- Difference: $A \setminus B = \{x \in \Omega : x \in A \wedge x \notin B\}$ (c.f. $A \wedge \not B$)

## Binary relations

- $A \subset B$ iff $x \in A \Rightarrow x \in B$ - (c.f. $A \implies B$)
- $A = B$ iff $x \in A \Leftrightarrow x \in B$ - (c.f. $A \Leftrightarrow B$)

# Functions

## Definition

A function $f : X \to Y$ is a relation from a set $X$ to a set $Y$ so that, for any $x \in X$, there exists a unique $y \in Y$ so that $y = f(x)$.

## Inverse of a function

A function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ if, for any $y \in Y$, there exists a unique $x \in X$ such that $f(x) = y$, and so that $f^{-1}(y) = x$.

# Sequences

## Sequences

A sequence $x_1, \ldots, x_t$ is written $\{x_i | i = 1, \ldots, t\}$ or $\{x_i\}$ for brevity

## Sums

A sum of a sequence of variables $\{x_i\}$ is defined as:

$$\sum_{i=1}^{t} x_i = x_1 + x_2 + \cdots + x_t.$$

The python notation `sum([f(x) for x in A])` corresponds to:

$$\sum_{x \in A} f(x)$$

## Products

A product is similarly defined at

$$\prod_{i=1}^{t} x_i = x_1 \times x_2 \times \cdots \times x_t.$$

# Events as sets

## The universe and random outcomes

- The $\Omega$ contains all events that can happen.
- When something happens, we observe an element $\omega \in \Omega$.

## Events in the universe

- An event is true if $\omega \in A$, and false if $\omega \notin A$.
- The negative event $\neg A = \Omega \setminus A$ is the set
- The possible events are a collection of subsets $\Sigma$ of $\Omega$ so that

(i) $\Omega \in \Sigma$, (ii) $A, B \in \Sigma \Rightarrow A \cup B in \Sigma$ (iii) $A \in \Sigma \Rightarrow \neg A \in \Sigma$

## Example: Traffic violation

- A car is moving with speed $\omega \in [0, \infty)$ in front of the speed camera.
- $A_0 = [0, 50]$: below the speed limit
- $A_1 = (50, 60]$: low fine
- $A_2 = (60, \infty]$: high fine
- $A_3 = (100, \infty)$: Suspension of license
- All combinations of the above events are interesting.

# Probability fundamentals

## Probability measure $P$

Probability can be seen as an area-like function assigning a likelihood to sets.

- $P : \Sigma \to [0, 1]$ gives the likelihood $P(A)$ of an event $A \in \Sigma$.
- $P(\Omega) = 1$
- For $A, B \subset \Omega$, if $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$.

## Marginalisation

If $A_1, \ldots, A_n \subset \Omega$ are a partition of $\Omega$

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i).$$

# Conditional probability

## Definition (Conditional probability)

The conditional probability of an event $A$ given an event $B$ is defined as

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$$

The above definition requires $P(B)$ to exist and be positive.

## Conditional probabilities as a collection of probabilities

More generally, we can define conditional probabilities as simply a collection of probability distributions:

$$\{P_\theta(A) \; |\theta \in \Theta\},$$

where $\Theta$ is an arbitrary set.

# The theorem of Bayes

**Theorem (Bayes's theorem)**

$$P(A|B) = \frac{P(B|A)}{P(B)}$$

# The theorem of Bayes

## Theorem (Bayes's theorem)

$$P(A|B) = \frac{P(B|A)}{P(B)}$$

## The general case

If $A_1, \ldots, A_n$ are a partition of $\Omega$, meaning that they are mutually exclusive events (i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$) such that one of them must be true (i.e. $\bigcup_{i=1}^{n} A_i = \Omega$), then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

and

$$P(A_j|B) = \frac{P(B|A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

# Independence

### Independent events

$A, B$ are independent iff $P(A \cap B) = P(A)P(B)$.

### Conditional independence

$A, B$ are conditionally independent given $C$ iff
$P(A \cap B|C) = P(A|C)P(B|C)$.

# Random variables

A random variable $f : \Omega \to \mathbb{R}$ is a real-value function measurable with respect to the underlying probability measure $P$, and we write $f \sim P$.

## The distribution of $f$

The probability that $f$ lies in some subset $A \subset \mathbb{R}$ is

$$P_f(A) \triangleq P(\{\omega \in \Omega : f(\omega) \in A\}).$$

## Independence

Two RVs $f, g$ are independent in the same way that events are independent:

$$P(f \in A \wedge g \in B) = P(f \in A)P(g \in B) = P_f(A)P_g(B).$$

In that sense, $f \sim P_f$ and $g \sim P_g$.

# Expectation

For any real-valued random variable $f : \Omega \to \mathbb{R}$, the expectation with respect to a probability measure $P$ is

$$\mathbb{E}_P(f) = \sum_{\omega \in \Omega} f(\omega) P(\omega).$$

When $\Omega$ is continuous, we can use a density $p$

$$\mathbb{E}_P(f) = \int_\Omega f(\omega) p(\omega) d\omega.$$

## Linearity of expectations

For any RVs $x, y$:

$$\mathbb{E}_P(x + y) = \mathbb{E}_P(x) + \mathbb{E}_P(y)$$

## Independence

If $x, y$ are independent RVs then $\mathbb{E}_P(xy) = \mathbb{E}(x)\,\mathbb{E}(y)$.

## Correlation

If $x, y$ are not correlated then $\mathbb{E}_P(xy) = \mathbb{E}(x)\,\mathbb{E}(y)$.

## IID (Independent and Identically Distributed) random variables

# Conditional expectation

The conditional expectation of a random variable $f : \Omega \to \mathbb{R}$, with respect to a probability measure $P$ conditioned on some event $B$ is simply

$$\mathbb{E}_P(f|B) = \sum_{\omega \in \Omega} f(\omega) P(\omega|B).$$

# Variance

For any real-valued random variable $f : \Omega \to \mathbb{R}$, the variance with respect to a probability measure $P$ is

$$\mathbb{V}_P(f) = \sum_{\omega \in \Omega} [f(\omega) - \mathbb{E}_P(f(\omega)]^2 P(\omega).$$

## Linearity of variance

If $f, g$ are uncorrelated RVs

$$\mathbb{V}_P(f + g) = \mathbb{V}_P(f) + \mathbb{V}_P(g).$$

## Variance products

If $f, g$ are independent RVs

$$\mathbb{V}_P(f + g) = \mathbb{E}_P(f)^2 \, \mathbb{V}_P(g) + \mathbb{E}_P(g)^2 \, \mathbb{V}_P(f) + \mathbb{V}_P(f) \, \mathbb{V}_P(g).$$

# Vector space $F$ axioms

- $(x + y) + z = x + (y + z)$, for all $x, y, z \in F$.
- $x + y = y + x$, for all $x, y \in F$.
- There is a zero element $0 \in F$ such that $x + 0 = 0$ for all $x \in F$.
- For all $x \in F$, there is an element $-x \in F$ so that $x + (-x) = 0$.
- $a(x + y) = ax + ay$, For any $a \in \mathbb{R}$, $x, y \in F$.
- $(a + b)x = ax + bx$, For any $a, b \in \mathbb{R}$, $x \in F$.

# The real vector space $F = \mathbb{R}^d$

For $a \in \mathbb{R}$ and $x, y \in F$,

- $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d)$
- $x + y = (x_1 + y_1, \ldots, x_d + y_d)$.
- $ax = (ax_1, \ldots, ax_d)$.
- $-x = (-1)x$.
- $0 = (0, \ldots, 0)$

# Linear operators

## Linear operator $A : F \to G$

- $A(x + y) = Ax + Ay$
- $A(ax) = a(Ax)$.

## Matrices in $\mathbb{R}^{n\times m}$.

A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ is a tabular array $\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix}$ Matrices can be seen as linear operators when used to multiply vectors.

# Multiplication operators

## Matrix multiplication

For $A \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{d \times m}$, the *ij*-th element of the result of the multiplication $AB$ is

$$(AB)_{i,j} = \sum_{k=1}^{d} A_{i,k} B_{k,j}.$$

so that $AB \in \mathbb{R}^{n \times m}$.

## Matrix-vector multiplication

A matrix $A \in \mathbb{R}^{n \times m}$ defines the following linear operator $A : \mathbb{R}^m \to \mathbb{R}^n$.

$$Ax = \left( \sum_{j=1}^{m} A_{i,j} x_j : i = 1, \ldots, n \right)$$

All vectors $x \in \mathbb{R}^m$ are equivalent to matrices in $\mathbb{R}^{m \times 1}$.

# Matrix inverses

## The identity matrix $I \in \mathbb{R}^{n \times n}$

- For this matrix, $I_{i,i} = 1$ and $I_{i,j} = 0$ when $j \neq i$.
- $Ix = x$ and $IA = A$.

## The inverse of a matrix $A \in \mathbb{R}^{n \times n}$

$A^{-1}$ is called the inverse of $A$ if

- $AA^{-1} = I$.
- or equivalently $A^{-1}A = I$.

## The pseudo-inverse of a matrix $A \in \mathbb{R}^{n \times m}$

- $\tilde{A}^{-1}$ is called the left pseudoinverse of $A$ if $\tilde{A}^{-1}A = I$.
- $\tilde{A}^{-1}$ is called the right pseudoinverse of $A$ if $A\tilde{A}^{-1} = I$.

# Derivatives

## Derivative

The derivative of a single-argument function is defined as:

$$\frac{d}{dx} f(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}.$$

$f$ must be absolutely continuous at $x$ for the derivative to exist.

## Subdifferential

For non-differential functions, we can sometimes define the set of all subderivatives:

$$\partial f(x) = [\lim_{\epsilon \to 0} \frac{f(x) - f(x-\epsilon)}{\epsilon}, \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}]$$

# Integrals

## Riemann integral

The Reimann integral is obtained by taking a horizontal discretisation of a function to the limit:

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{t=1}^{n} f(x_t)\frac{b-a}{n}, \qquad x_t = a + (t-1)\cdot\frac{b-a}{n}$$

## Lebesgue integral

This integral is obtained by taking a vertical discretisation of a function to the limit. Let $\lambda$ be the Lebesgue measure (i.e. area) of a set. Then:

$$\int_X f(x)d\lambda(x) = \lim_{n\to\infty} \sum_{t=1}^{n} y_t \lambda(S_t),$$

$S_t = \{x : f(x) \in (y_{t-1}, y_t\}$, $y_0 = -\infty$, $y_n = \sup_x f(x)$.

# Fundamental theorem of calculus

$$f(x) = \frac{d}{dx}\int_a^x f(t)dt$$

If $\frac{d}{dx}F = f$ then its integral from $a$ to $b$ is:

$$\int_a^b f(x)dx = F(b) - F(a),$$

# Multivariate Functions

We consider functions operating in multi-dimensional Euclidean spaces.

$f : \mathbb{R}^n \to \mathbb{R}$.

- Any $x \in \mathbb{R}^n$ is $x = (x_1, \dots, x_n)$, with $x_i \in \mathbb{R}$.
- We write $f(x)$ instead of $f(x_1, \dots, x_n)$.

$f : \mathbb{R}^n \to \mathbb{R}^m$.

- If $y = f(x)$ then $y_i$ is the $i$-th component of $y \in \mathbb{R}^m$.
- Can be seen as $m$ functions $f_i : \mathbb{R}^n \to \mathbb{R}$, with $y_i = f_i(x)$.

# Derivatives in many dimensions

## Partial derivative

The partial derivative of $f$ with respect to its $i$-th argument is: $\frac{\partial}{\partial x_i} f(x)$, where we see all $x_j$ with $j \neq i$ as fixed.

## Gradient of $f$

This is the vector of all its partial derivatives:

$$\nabla_x f(x) = \left( \frac{\partial}{\partial x_1} f(x) \cdots \frac{\partial}{\partial x_i} f(x) \cdots \frac{\partial}{\partial x_n} f(x) \right)^\top$$

## Directional derivative

$$D_\delta f(x) = \lim_{\epsilon \to 0} \frac{f(x + \epsilon\delta) - f(x)}{\epsilon}.$$