MASTER IMALIS - ENS PSL

Training in Mathematics and Statistics

SEPTEMBER 2020

Benoît Perez-Lamarque - benoit.perez@ens.psl.eu

Antoine Sicard - antoine.sicard@ens.psl.eu

Planning

Lecture 1: Few revisions

1.1 Sets

1.1.1 Common sets

By convention, the following symbols are reserved for the most common sets of numbers:

$\varnothing$ – empty set;

$\mathbb{N}$ – natural numbers, $\mathbb{N} = \{0, 1, 2, ...\}$;

$\mathbb{Z}$ – integers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$;

$\mathbb{Q}$ – rational numbers (from quotient), $\mathbb{Q} = \left\{\frac{p}{q}, p \in \mathbb{Z}, q \in \mathbb{N}^*\right\}$;

$\mathbb{R}$ – real numbers, $\mathbb{R} = \{a_1 a_2 \ldots a_p.a_{p+1} ..., \forall i \in \mathbb{N}, a_i \in \mathbb{N}, p \in \mathbb{N}^*\}$;

$\mathbb{C}$ – complex numbers, $\mathbb{C} = \{\alpha + i\beta, (\alpha, \beta) \in \mathbb{R}^2\}$. α (resp. β) is referred to as the real part (resp. the imaginary part), and the imaginary unit i is defined by its property $i^2 = -1$.

1.1.2 Product of sets

Let E and F be two sets:

— $E \times F = \{(x, y), x \in E, y \in F\}$;

— $E \times E = E^2$ is the set of all couples of E;

— $E \times \ldots \times E = E^n$ is the set of n-tuple of E.

1.2 Functional analysis

1.2.1 Asymptotic notation

Let f and g be two functions in the neighbourhood of a, such as g is not equal to 0 in the neighbourhood of a.
The function f is **negligible** with respect to g in the neighbourhood of a, if $\lim_{x \to a} \frac{f(x)}{g(x)} = 0$, and f is denoted: $f = \circ(g)$ (called *"little-o"*).

In other words, $f(x)/g(x)$ tends to zero as x tends to a and the limit of f/g at a is zero.

1.2.2 Continuity

A function $f : E \to \mathbb{R}$ is **continuous** at $x_0 \in E$ if $\lim_{x \to x_0} f(x) = f(x_0)$.

To go further, f is continuous at x_0 if, for $\epsilon \to 0$, $f(x_0 + \epsilon) = f(x_0) + \circ(1)$.

1.2.3 Derivability

A function f is **differentiable** at $x_0 \in E$ if $\frac{f(x) - f(x_0)}{x - x_0}$ has a limit when $x \to x_0$. This limit is referred to as the **derivative** of f at x_0, denoted $f'(x_0)$.

Other notation: $f' = \frac{df}{dx}$.

If $f(x, y)$ is a function of several variables (x and y), the **partial derivatives** of f are the derivatives of f with respect to one of its variables (either x or y), denoted:

$$\frac{\partial f(x, y)}{\partial x} \text{ or } \frac{\partial f(x, y)}{\partial y}$$

Common derivative:

Let $c \in \mathbb{R}$ be a constant , $\forall x \in \mathbb{R}$:

$f(x) = c$ has for derivative $f'(x) = 0$;

$f(x) = cx$ has for derivative $f'(x) = c$;

$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}, f(x) = cx^n$ has for derivative $f'(x) = cnx^{n-1}$;

$\forall x \in \mathbb{R}^*, \forall \alpha \in \mathbb{Z}, f(x) = cx^\alpha$ has for derivative $f'(x) = c\alpha x^{\alpha-1}$ (and so $f(x) = x^{-1} = \frac{1}{x}$ has for derivative $\frac{-1}{x^2}$);

$\forall x \in \mathbb{R}_+^*, \forall \alpha \in \mathbb{R}, f(x) = cx^\alpha$ has for derivative $f'(x) = c\alpha x^{\alpha-1}$ (and so $f(x) = x^{1/2} = \sqrt{x}$ has for derivative $\frac{1}{2\sqrt{x}}$);

$f(x) = e^{cx}$ has for derivative $f'(x) = ce^{cx}$;

$\forall x \in \mathbb{R}_+^*, f(x) = \ln(x)$ has for derivative $f'(x) = \frac{1}{x}$.

$\forall$ a a constant $\in \mathbb{R}_+^*, \forall x \in \mathbb{R}, f(x) = a^x$ has for derivative $f'(x) = a^x \ln(a)$.

Operations on derivative: Let $c \in \mathbb{R}$ be a constant and f and g two functions :

— scalar multiplication: $(cf)' = cf'$;

— sum of two functions: $(f+g)' = f' + g'$;

— product of two functions: $(fg)' = f'g + fg'$;

— function composition: $(f \circ g)' = g' \; f' \circ g$;

— inverse function: $\left(\frac{1}{f}\right)' = -\left(\frac{-f'}{f^2}\right)$

— quotient of two functions: $\left(\frac{f}{g}\right)' = \left(\frac{f'g - fg'}{g^2}\right)$.

1.2.4 Bijectivity

A function $f : E \to F$ is **injective** for all a and b in E, if and only if (iif), $f(a) = f(b)$ implies $a = b$.

A function $f : E \to F$ is **surjective**, iif for every element $y \in F$, there is at least one element $x \in E$ such that $f(x) = y$.

A function $f : E \to F$ is **bijective** (or one-to-one correspondence), iif f is injective and surjective at the same time, i.e. every $y \in F$ has a unique counterimage with f:

$$\forall y \in F, \exists! x \in E, f(x) = y$$

If f is bijective, one can define a function g that associates to every $y \in F$ its counterimage with f. It verifies $g \circ f = IdE$ and $f \circ g = IdF$, where IdE and IdF represent the identity function: $\forall x \in E, \; g \circ f(x) = x$ and $\forall y \in F, \; f \circ g(y) = y$).

g is called **inverse function** of f, $g = f^{-1}$.

1.2.5 Differential equation

A **differential equation** is an equation involving an unknown function f and at least one of its derivatives (f', f'',...). If the unknown function f only involves derivatives with respect to one variable, then the differential equation is called an **ordinary differential equation** (ODE).

For example, $\forall (a, b) \in \mathbb{R}$, the differential equation of first order $f' + af = b$ has for set of solutions the functions defined by:

$$\forall \lambda \in \mathbb{R}, \; \forall x \in \mathbb{R}, \; f(x) = \lambda e^{-ax} + \frac{b}{a}$$

The value of the arbitrary constant λ can be found by assuming particular conditions (e.g. initial conditions).

If the unknown function involves derivatives with respect to two or more variables (x, y, ...), then the differential equation is called a **partial differential equation** (PDE).

1.3 Matrix

1.3.1 Definitions

— A **matrix** is any rectangular array of numbers. If the array has n rows and m columns, then it is an $n \times m$ matrix, denoted $A_{n,m}$. One dimensional matrices are called row vectors for a $1 \times m$ matrix or column vectors for a $n \times 1$ matrix. One uses the notation $(a_{i,j})$ to refer to the number in the $i - th$ row and $j - th$ column. If $n = m$, $A_{n,m} = A_{n,n} = A_n$ is called a **square matrix**.

— The zero matrix or null matrix is a matrix with all its elements equal to zero, denoted $0_{n,m}$.

— The **identity matrix** is a square matrix with ones on the main diagonal and zeros elsewhere, called I_n. The identity matrix is neutral with regard to products: $\forall A_n$, $A \times I_n = I_n \times A = A$.

— The **trace**, called tr(A), of a square matrix A is the sum of its diagonal elements.

1.3.2 Matrix operation

— The **transpose** of a matrix flips a matrix $A = [a_{i,j}]$ over its diagonal: it switches the row and column indices of the matrix and gives another matrix denoted as ${}^tA$ (also called A', A^{tr}, or A^T): ${}^tA = [a_{j,i}]$.

— The matrix addition is the operation of adding two matrices of the same dimensions, $A_{n,m}$ and $B_{n,m}$, by adding the corresponding elements together.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix}$$

— The multiplication by a scalar λ: $\lambda(a_{i,j}) = (\lambda a_{i,j})$.

$$\lambda \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}$$

— The matrix product : we can only multiply two matrices together if the number of columns of the first matrix equals the number of rows of the second matrix.
Let $A_{n,m}$ and $B_{m,p}$ be two matrices: $A_{n,m}B_{m,p}$ exists but $B_{m,p}A_{n,m}$ does not exist if $n \neq p$.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

Some properties on the matrix product:

Let A, B, and C be three matrices (such that their products exist), and μ and λ two scalars :

i) $AB \neq BA$ in general: the matrix product is not commutative;

ii) $\lambda(AB) = (\lambda A)B = A(\lambda B)$: the matrix product is associative;

iii) ${}^t(AB) = {}^tB\,{}^tA$

iv) $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$.

v) $AB = 0$ does not imply $A = 0$ or $B = 0$. Moreover, $AC = BC$ does not imply $A = B$.

1.3.3 Determinant of a square matrix

The **determinant** is a value that can be computed from the elements of a square matrix A_n, denoted $\det(A) = |A|$.

For $n = 2$, if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

If $n > 2$, the determinant is defined recursively using the Laplace formula with regard to a row or a column and using cofactors. For example, if $n = 3$:

$$\det(A) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} \oslash & \oslash & \oslash \\ \oslash & e & f \\ \oslash & h & i \end{vmatrix} - b \begin{vmatrix} \oslash & \oslash & \oslash \\ d & \oslash & f \\ g & \oslash & i \end{vmatrix} + c \begin{vmatrix} \oslash & \oslash & \oslash \\ d & e & \oslash \\ g & h & \oslash \end{vmatrix}$$

$$= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} = a(ei - hf) - b(di - gf) + c(dh - ge)$$

For a triangular matrix, its determinant is the product of its diagonal elements.

1.4 Counting

The **cardinality** of a set E, called $card(E)$ is the number of elements of the set E.

$\forall n \in \mathbb{N}$, the **number of permutations** of the n elements, denoted $\boldsymbol{n!}$ (and called *"n-factorial"*), is defined as:

$$n! = \begin{cases} 1 \times 2 \times \ldots \times (n-1) \times n & \text{if } n > 0 \\ 1 & \text{if } n = 0. \end{cases}$$

An **arrangement** is an ordered subset of k elements among n. The **number of arrangement** A_n^k of k elements among n is defined as:

$$A_n^k = \frac{n!}{(n-k)!}$$

A **combination** is a (unordered) subset of k elements among n. The **number of combination** C_n^k is defined as:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

1.5 Discrete probability

1.5.1 Probability space

Let's assume a randomized experiment (when the outcome is not deterministic, but the probability of each event is known) defined by a **probability space** $\boldsymbol{(\Omega, P)}$:

— Ω is the set of all possible outcomes, called **sample space**.

— P is the **probability distribution** associated to the outcomes of the experiment. P verifies:

$$\begin{cases} \forall x \in \Omega, P(x) \in [0,1] \\ \sum_{x \in \Omega} P(x) = 1 \end{cases}$$

An **event** E is a subset of Ω and verifies: $P(E) = \sum_{x \in E} P(x)$

If all events of Ω are elementary events (i.e. all events are equiprobable), then $\forall E \in \Omega$:

$$P(E) = \frac{card(E)}{card(\Omega)}$$

Let (Ω, P) be a probability space and A and B two events from this space:

(i) $P(A) \in [0,1]$;

(ii) $P(\varnothing) = 0$ and $P(\Omega) = 1$;

(iii) The **complementary event** of A, denoted $\overline{A}$ or A^c, verifies: $P(\overline{A}) = 1 - P(A)$;

(iv) The probability of having A and B is denoted $P(A \cap B)$;

(v) The probability of having A or B is: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

(vi) The events A and B are **incompatible** iif $A \cap B = \varnothing$. Then, $P(A \cup B) = P(A) + P(B)$.

1.5.2 Conditional probability and independence

A. Conditional probability

Given a probability space (Ω, P) and two events A and B with $P(B) \neq 0$. The conditional probability of A given B, denoted $P(A|B)$ or $P_B(A)$, is defined by:

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

Consequently, $P(A \cap B) = P(A|B)P(B)$

One can deduce:

(i) the **Bayes' theorem**:

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

(ii) the **law of total probability**:

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A \cap \overline{B})P(\overline{B})$$

B. Independence

Two events A and B are independent iif $P(A \cap B) = P(A)P(B)$.

Similarly, if $P(B) \neq 0$, A and B are independent iif $P(A|B) = P(A)$.

1.6 Taylor series

The Taylor series of a function is a series expansion of the function in the neighbourhood of a point. For example, the Taylor series of a function $f(x)$ around a certain value a is

$$f(x) = f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)(x-a)^2}{2!} + \frac{f'''(a)(x-a)^3}{3!} + ... + \frac{f^n(a)(x-a)^n}{n!}$$

The Taylor series is very useful to approximate a complex function around a certain point and is often used in the analysis of non-linear biological system.

1.7 Other revisions

— $\forall\ (a,b) \in \mathbb{R}^2$, $(a+b)^2 = a^2 + 2ab + b^2$, and $a^2 - b^2 = (a-b)(a+b)$.

— $\forall\ (a_1, \ldots, a_n) \in \mathbb{R}^n$, $(a_1 + \ldots + a_n)^2 = \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j$

— Two vectors $v_1 = (x,y)$ and $v_2 = (x', y')$ are collinear if $\exists\ a \in \mathbb{R}, v_1 = a v_2$ that is to say, $xy' + yx' = 0$;

— $\forall\ \theta \in \mathbb{R}$, $\cos(\theta) + i\sin(\theta) = e^{i\theta}$.

— $f : \mathbb{R} \to \mathbb{R}$ is an even function iif $\forall\ x \in \mathbb{R},\ f(-x) = f(x)$.

— $f : \mathbb{R} \to \mathbb{R}$ is an odd function iif $\forall\ x \in \mathbb{R},\ f(-x) = -f(x)$.