Mathematics
Academic year 2023-2024
Teacher: P. Gibilisco

Lecture 1 – Monday, September 4, 2023 (14:00-16:15)

The beginner ... should not be discouraged if ... he finds that he does not have the prerequisites for reading the prerequisites.
P.Halmos

Introduction to the course.

An example on the applications of mathematics in economics: risk aversion and concavity.

List of the notions required in advance to follow the course.

Deterministic and random phenomena.

The "law of chance" as an oxymoron.

Axiomatization of geometry and probability: Euclid and Kolmogorov.

The ideas of Kolmogorov: events are represented by sets; probability is a normed measure on them.

Probability spaces and their properties (finitely additive case).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

No proposition is in itself either probable or improbable, just as no place can be intrinsically distant ...
J.M.Keynes

Conditional probability, pairs of independent events ($A \perp B$ means A and B are independent). Remark: $P(A) = P(A|\Omega)$.

The two pillars of probability theory: on one side analysis and measure theory; on the other side gambling situations, coin-tossing.

The geometric series and its sum.

$$1 + q + q^2 + q^3 + \cdots = \frac{1}{1-q} \quad \text{if} \qquad |q| < 1$$

Lecture 2 – Tuesday, September 5, 2023 (14:00-16:15)

... the fact is that there is nothing as dreamy and poetic, nothing as radical, subversive, and psychedelic, as mathematics ... Mathematics is the purest of the arts, as well as the most misunderstood.
P.Lockhart

Exercise: prove that $A \perp B \Rightarrow A \perp B^c$

Bayes formula.

Partitions.

Exercise: an urn contains w white balls and r red balls. Choose a ball in the urn and leave it out (without looking at its colour). Choose a second ball. Which is the probability that the second ball is white? (The answer will explains the queue paradox).

Taylor polynomial and Taylor series. Example:

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad (|x| < 1)$$

Taylor polynomial and Taylor series. Example:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \in \mathbb{R})$$

Pairwise independence and independence.

A counterexample: pairwise independence does not imply independence. Exercise: prove that

$$p(A|B) + p(A^c|B) = 1,$$

while (in general)

$$p(A|B) + p(A|B^c) \neq 1.$$

(Conditional probability is a probability measure only with respect to the first argument!)

Lecture 3 – Wednesday, September 6, 2023 (14:00-16:15)

No one shall expel us from the paradise which Cantor has created for us.
D. Hilbert

Countable and uncountable sets. Cardinality of $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$.

Counting measure on $\mathbb{N}$ is not continuous.

σ–algebras and σ–additivity for a probability measure.

Probability spaces and their properties (σ–additive case).

Continuity of a probability measure. Equivalence of continuity from above and σ–additivity (no proof!)

Counting subsets of a set (n choose k). Newton binomial formula. The cardinality of the power set: $2^n = \sum_{k=0}^{n} \binom{n}{k}$.

Coin flipping and the Bernoulli process. The probability to have k successes flipping a coin n times:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Lecture 4 – Thursday, September 7, 2023 (14:00-16:15)

Examples about our probabilistic intuition.

Probability test: 1) the birthday problem; 2) the false positive problem; 3) the Monty-Hall problem; 4) the queue problem.

Solution of the above problems.

Image of a function.

Random variables: discrete case

Distribution of a random variable.

Exercise. A fair die is rolled two times. Let us denote by X the first result and by Y the second result. Let $U := X - Y$. Which is the distribution of the random variable U?

Non-negative series with sum equal to one and discrete distributions.

Bernoulli process.

The most important discrete r.v.: Bernoulli, binomial, geometric and Poisson distribution.

Lecture 5 – Friday, September 8, 2023 (14:00-16:15)

Independence for discrete random variables.

Expected value for discrete r.v.

Linearity and positivity for the expected value.

How the Bernoulli, Binomial, Geometric and Poisson distributions arise flipping a coin (Bernoulli process).

$$\lim_{n\to+\infty}\left(1+\frac{c}{n}\right)^n = e^c$$

Convergence of the Bernoulli to the Poisson.

Mean value for the Bernoulli and binomial distribution.

A counterexample. Consider a box where there are three balls with the numbers $-1, 0, 1$. Choose randomly a ball and let X be the random variable which represent the result so that:

$$P(X=-1) = P(X=0) = P(X=1) = \frac{1}{3}.$$

Prove that $X \not\perp X^2$.

Lecture 6 – Monday, September 11, 2023 (14:00-16:15)

$$\mathbb{E}(g(X)) = \sum_k g(x_k)P(X=x_k)$$

Moments of a r.v.. Variance.

Covariance and its properties.

Standard deviation.

Correlation coefficient and scale invariance.

$$X \perp Y \qquad \Rightarrow \qquad \mathrm{E}(XY) = \mathrm{E}(X)\cdot\mathrm{E}(Y)$$

$$X \perp Y \qquad \Rightarrow \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$X \perp Y \qquad \Rightarrow \qquad \mathrm{Cov}(X,Y) = 0$$

A counterexample. Consider a box where there are three balls with the numbers $-1, 0, 1$. Choose randomly a ball and let X be the random variable which represent the result so that:

$$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}.$$

Prove that $X \not\perp X^2$ while $\mathrm{Cov}(X, X^2) = 0$ (this is true in general for X and X^2 when X has a distribution which is symmetric w.r.t. to 0).

This prove that:

$$X \perp Y \qquad \not\Leftarrow \qquad \mathrm{Cov}(X, Y) = 0$$

Variance for the Bernoulli and Binomial distribution.

Cumulative distribution function.

Continuous random variables ($P(X = x) = 0 \, \forall x \in \mathbb{R}$, continuity of F_X).

Densities and absolutely continuous random variable.

The uniform density.

Exponential density.

Lecture 7 – Tuesday, September 11, 2023 (14:00-16:15)

Moments and variance for absolutely continuous random variables.

Mean value and variance for the uniform distribution.

Mean value and variance for the exponential distribution.

Cumulative distribution function for uniform and exponential distribution.

A mathematician is someone to whom the facts that

$$2 + 2 = 4$$

and

$$\int_{\mathbb{R}} e^{-x^2} \, dx = \sqrt{\pi}$$

are equally obvious.
Lord Kelvin

If X is an a.c. r.v. with density f_X then

$$F'_X = f_X$$

Functions whose antiderivative cannot be expressed as an elementary function.

$$\int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi} \qquad \text{(No proof)}$$

The Gaussian density

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Lecture 8 – Wednesday, September 13, 2023 (14:00-16:15)

Density of $aX + b$ from the density of X.

$$\int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2}}\,\mathrm{d}x = \sqrt{2\pi} \qquad \text{Integration by parts}$$

Mean and variance for the gaussian density.

The Gaussian density: the general case ($X \sim \mathcal{N}(\mu, \sigma^2)$ means $X = \sigma Z + \mu$ where $Z \sim \mathcal{N}(0, 1)$.)

Convergence in probability.

Chebyshev's inequality.

The (weak) law of large numbers.

Lecture 9 – Monday, September 18, 2023 (14:00-16:15)

Written in five years, may it last as many thousands.
G. Cardano at the end of Ars Magna

Some properties of power series: the radius of convergence.

Mac Laurin series for the trigonometric functions.

Introduction to complex numbers.

Exercise: calculate

$$Re\left(\frac{1}{2+3i}\right) = \cdots$$

History of $0, 1, i, e, \pi$.

Complex exponential as a power series.

$$e^{i\theta} = \cos\theta + i\sin\theta$$

The Euler formula.

$$e^{i\pi} + 1 = 0$$

Lecture 10 – Tuesday, September 19, 2023 (14:00-16:15)

Students don't need a perfect teacher. Students need a happy teacher, who's gonna make them excited to come to school and grow a love for learning.
R. Feynman

The Fundamental theorem of Algebra.
Standardized random variables

$$X^* := \frac{X - \mathbb{E}(X)}{\sigma(X)}$$

$$\rho(X^*, Y^*) = \rho(X, Y)$$

Derivatives of power series, especially the geometric series.
Complex random variables and their expectations.
Quiz
If $A = \emptyset, \Omega$ then $A \perp B$ for any event B.
Quiz
A constant random variable is independent from any random variable.
Remark: $X \perp Y$ implies $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$.
The characteristic function, first properties.

- $\varphi_X(0) = 1$
- $\varphi_{\lambda X}(t) = \varphi_X(\lambda t)$
- $X \perp Y \qquad \Rightarrow \qquad \varphi_{X+Y}(t) = \varphi_X(t)\,\varphi_Y(t)$

The inversion theorem (no proof!):

X, Y have the same distribution iff $\varphi_X(t) = \varphi_Y(t) \qquad \forall t \in \mathbb{R}$

$$X \sim \text{Poisson}(\lambda) \Longrightarrow \varphi_X(t) = \exp(\lambda(e^{it} - 1))$$

Using the characteristic function prove that

$$X \perp Y, \quad X \sim \text{Poisson}(\lambda), \quad Y \sim \text{Poisson}(\mu) \Longrightarrow X + Y \sim \text{Poisson}(\lambda + \mu)$$

If $X_n \sim B(n, p)$ then $\varphi_X(t) = (pe^{it} + q)^n$. Using this result prove that the sum of independent Bernoulli variables is binomial.

Lecture 11 – Wednesday, September 20, 2023 (14:00-16:15)

Partial derivatives. Derivation under the integral sign.

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$$

Moments and the characteristic function. The Taylor formula for $\varphi_X(\cdot)$.

- Convergence in law (in distribution).
- Example: if X_n and X are random variables whose values are natural numbers then convergence in law is equivalent to (no proof)

$$P(X_n = k) \to P(X = k) \qquad \forall k \in \mathbb{N}$$

- Example: if $X_n \sim B(n, \frac{\lambda}{n})$ and $X \sim \text{Poisson}(\lambda)$ then $X_n \Longrightarrow X$ in law.
- Continuity theorem (no proof!):

$$X_n \Longrightarrow X \quad \text{in law} \qquad \text{iff} \qquad \varphi_{X_n}(t) \to \varphi_X(t) \qquad \forall t \in \mathbb{R}$$

- Example: prove that if $X_n \sim B(n, \frac{\lambda}{n})$ and $X \sim \text{Poisson}(\lambda)$ then $X_n \Longrightarrow X$ in law using the characteristic function and the continuity theorem.
- Characteristic function of the gaussian: if $X \sim \mathcal{N}(0, 1)$ then $\phi_X(t) = e^{-\frac{t^2}{2}}$. (No proof!)
- CLT. $X_1, ..., X_n, ...$ i.i.d. random variables with finite mean and variance. Then

$$S_n^* \longrightarrow \mathcal{N}(0, 1) \qquad \text{in law}$$

- Meaning of the CLT: the sum of (infinitely) many, small, independent effects is (approximately) normal (gaussian).

- CLT reformulated. Let $\Phi(t)$ the c.d.f. of the standard gaussian distribution.

  $X_1, ..., X_n, ...$ i.i.d. random variables with finite mean μ and variance σ^2. Then

$$P(X_1 + \cdots + X_n \leq x) = P\left(S_n^* \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right) \to \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Counterexample: S random sign, $X \sim \mathcal{N}(0,1)$, $S \perp X$ then $Y = SX \sim \mathcal{N}(0,1)$. Moreover $P(X + Y = 0) = \frac{1}{2}$ therefore $X + Y$ is not gaussian.

Lecture 12 – Thursday, September 21, 2023 (14:00-16:15)

The Gamma function

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} \mathrm{d}x \qquad \alpha > 0$$

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha) \qquad \Gamma(1) = 1 \qquad \Gamma(n) = (n-1)!$$

$X \sim \Gamma(\alpha, \lambda)$ (where $\alpha, \lambda > 0$) if

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \cdot 1_{(0,\infty)}(x)$$

If

$$c \cdot x^{\alpha-1} e^{-\lambda x} \cdot 1_{(0,\infty)}(x)$$

is a density then

$$c = \frac{\lambda^\alpha}{\Gamma(\alpha)}$$

If X has density f_X then X^2 has density

$$f_{X^2}(x) = \frac{1}{2\sqrt{x}}(f(\sqrt{x}) + f(\sqrt{-x})) \cdot 1_{(0,\infty)}(x)$$

If $X \sim \mathcal{N}(0,1)$ then $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.
This implies

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

- Moments of the Gamma distribution. If $X \sim \Gamma(\alpha, \lambda)$ then

$$\mathbb{E}(X^n) = \frac{1}{\lambda^n} \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}$$

$$\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$$

If $X \sim \Gamma(\alpha, \lambda)$ then

- The characteristic function (no proof):

$$\varphi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^{\alpha}$$

- If $X \sim \Gamma(\alpha, \lambda)$, $Y \sim \Gamma(\beta, \lambda)$ and $X \perp Y$ then $X + Y \sim \Gamma(\alpha + \beta, \lambda)$.
- The chi-squared distribution with n degrees of freedom. If $X \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ then we say that $X \sim \chi^2(n)$.
- If $X_1, X_2, ..., X_n$ are independent standard gaussian then $X_1^2 + X_2^2 + \cdots X_n^2 \sim \chi^2(n)$.

Lecture 13 & 14 – Tuesday, September 26, 2023 (11-13 & 14:00-16:15)

Real vector spaces and scalar products.

Examples: $\mathbb{R}^n$, real polynomials, continuous functions on the unit interval, random variables on a probability spaces.

Exercise: prove that $E(X)$ is the constant "nearest" to the random variable X w.r.t. the distance associated to scalar product $\langle X, Y \rangle = E(XY)$.

Cauchy-Schwartz inequality in the plane, for functions, for random variables. Application: $|\rho(X, Y)| \leq 1$.

- Random vectors, discrete case.
- Marginals.

Exercise. Consider a urn with w white balls and b black balls. You can extract balls with and without replacement.

Replacement case. Define

$$X_1 = \begin{cases} 1 & \text{if the first ball is white} \\ 0 & \text{otherwise} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{if the second ball is white} \\ 0 & \text{otherwise} \end{cases}$$

No replacement case. Define

$$\tilde{X}_1 = \begin{cases} 1 & \text{if the first ball is white} \\ 0 & \text{otherwise} \end{cases} \qquad \tilde{X}_2 = \begin{cases} 1 & \text{if the second ball is white} \\ 0 & \text{otherwise} \end{cases}$$

Prove that the joint distributions of the random vectors (X_1, X_2), $(\tilde{X}_1, \tilde{X}_2)$ are different while the marginal distributions are equal.

One shouldn't never integrate in public.
R. Feynman

Introduction to multiple integrals. The Fubini theorem.
Let $Q = [1, 2] \times [1, 3]$ and $f(x, y) = x^3 \exp(yx^2)$. Show that

$$\int\int_Q f(x, y) = \frac{1}{6}[e^{12} - e^3] - \frac{1}{2}[e^4 - e]$$

Show that to calculate the integral

$$\int\int_Q f(x, y) = \frac{1}{6}[e^{12} - e^3] - \frac{1}{2}[e^4 - e]$$

the order in which we choose the variables matters!!!!
Normal domains. Prove that if

$$A = \{(x, y) \in \mathbb{R}^2 | x \in [-1, 1] \quad -1 + x^2 \leq y \leq \sqrt{1 - x^2}\}$$

then

$$\int\int_A xy \, \mathrm{d}x\mathrm{d}y = 0$$

Lecture 15 – Wednesday, September 27, 2023 (14:00-16:15)

Random vectors: absolutely continuous case.
Proposition $X \perp Y \iff f_{X,Y} = f_X \cdot f_Y$
Exercise. Suppose that the random vector (X, Y) has the joint density $f_{X,Y}(x, y) = g(x)h(y)$. Then $X \perp Y$.

- Gaussian vectors: the standard case.
- The density of the standard case:

$$\frac{1}{2\pi}\exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

- Gaussian vectors: the general case.

- The density of a bivariate gaussian (no proof!):

$$f_{X_1,X_2}(x,y) =$$
$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\right)\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\Bigg)$$

- If X, Y are jointly gaussian random variables then $X \perp Y$ is equivalent to $\mathrm{Cov}(X,Y) = 0$.
- A linear combination of jointly gaussian random variables is gaussian.
- Remember! S random sign, $X \sim \mathcal{N}(0,1)$, $S \perp X$ then $Y = SX \sim \mathcal{N}(0,1)$. Moreover $P(X+Y=0) = \frac{1}{2}$ therefore $X+Y$ is not gaussian.

  Therefore X and Y are gaussian but not jointly gaussian.

Exercise

Let us consider a Gaussian vector

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 2 \\ a \end{pmatrix}\begin{pmatrix} 3 & b \\ -1 & 1 \end{pmatrix}\right)$$

such that $\mathrm{E}(XY) = 2$.
i) Calculate a e b.
ii) For which values of c and d are the random variables $dX - cY$ and X independents?

Exercise

Let us consider a standard Gaussian vector

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

Then

$$P(Z_1 > 0, Z_2 > 0) = \frac{1}{4}, \qquad P(Z_1 > Z_2) = \frac{1}{2}.$$

Lecture 16,17,18,19 – October 2,3,4,5, 2023 (14:00-16:15)

- Search for the random variable $\mu(Y)$ nearest to X (w.r.t. the L^2 distance).
- Remember $\mathbb{E}(AB) = \sum_{k,j} a_k b_j P((A = a_k) \cap (B = b_j))$.
- Conditional expectation for discrete random variables.
- Conditional expectation for discrete random variables (linearity, positivity, constants).

$$\mathbb{E}(g(Y)X|Y) = g(Y)\mathbb{E}(X|Y)$$

- $$\mathbb{E}(g(Y)\mathbb{E}(X|Y)) = \mathbb{E}(g(Y)X)$$

  (Namely: any r.v. $g(Y)$ is orthogonal to the r.v. $(X - \mathbb{E}(X|Y))$.
- Law of iterated expectations

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$$

- $$\mathbb{E}((X-g(Y))^2) = \mathbb{E}((X-\mathbb{E}(X|Y))^2)+\mathbb{E}((g(Y)-\mathbb{E}(X|Y))^2) \geq \mathbb{E}((X-\mathbb{E}(X|Y))^2)$$

  Namely: $\mathbb{E}(X|Y)$ is the random variable $\mu(Y)$ nearest to X (w.r.t. the L^2 distance).

$$\mathbb{E}(g(Y)|Y) = g(Y)$$
$$X \perp Y \Longrightarrow \mathbb{E}(X|Y) = \mathbb{E}(X)$$

Some counterexamples.

$$\mathbb{E}(X|Y) = \mathbb{E}(X) \qquad \Longrightarrow \qquad \mathrm{Cov}(X, Y) = 0$$

The viceversa it is not true (use a symmetric distribution X such that $\mathrm{Cov}(X^2, X) = 0$ but $\mathbb{E}(X^2|X) = X^2 \neq \mathbb{E}(X^2)$).

- S random sign, $X \sim \mathcal{N}(0,1)$, $S \perp X$ and $Y := SX$ (we proved $Y \sim \mathcal{N}(0,1)$ and $Y \not\perp X$). Then

$$\mathbb{E}(SX|X) = X\mathbb{E}(S|X) = X\mathbb{E}(S) = 0$$

and

$$\mathbb{E}(SX) = \mathbb{E}(Y) = 0$$

so that $\mathbb{E}(SX|X) = \mathbb{E}(SX)$ but $SX \not\perp X$.

- Conclusion: we have that

$$X \perp Y \quad \Longrightarrow \quad \mathbb{E}(X|Y) = \mathbb{E}(X) \quad \Longrightarrow \quad \mathrm{Cov}(X,Y) = 0$$

while

$$X \perp Y \quad \not\Longleftarrow \quad \mathbb{E}(X|Y) = \mathbb{E}(X) \quad \not\Longleftarrow \quad \mathrm{Cov}(X,Y) = 0$$

Conditional expectation for jointly gaussian random variable in some steps.

- If $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ then $(X - \alpha Y)$ and Y are L^2 orthogonal if and only if

$$\alpha = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}.$$

- Let X, Y be jointly gaussian, $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ and let

$$Z := X - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} \cdot Y.$$

Then Z and Y are independent (and $\mathbb{E}(Z) = 0$).

- X, Y jointly gaussian and $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ implies

$$\mathbb{E}(X|Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}Y} \cdot Y = \rho\frac{\sigma_X}{\sigma_Y} \cdot Y$$

- Let $k, h, \alpha \in \mathbb{R}$. Then

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X+k, Y+h)$$
$$\mathbb{E}(U|Y) = \mathbb{E}(U|Y+\alpha)$$

- X, Y jointly gaussian implies

$$\mathbb{E}(X|Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}Y}(Y - \mathbb{E}(Y)) + \mathbb{E}(X) = \rho\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y) + \mu_X$$

  (Once you calculate a conditional expectation check the law of iterated expectation.)

- The conditional variance

$$\mathrm{Var}(X|Y) = \mathbb{E}((X - \mathbb{E}(X|Y))^2|Y)$$

$$\mathrm{Var}(X|Y) = \mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2$$

- The law of total variance

$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathbb{E}(X|Y))$$

Vector spaces, linear transformations, scalar products. Linear independence, basis.

Kernel of a linear transformation.

Sets of generators in a vector space. Characterization of the basis as: 1) maximal sets of linearly independent vectors; 2) minimal sets of generators.

Existence of a basis for vector spaces.

Dimension of vector spaces as cardinality of basis.

Matrices and their operations. Representation of linear transformation by matrices.

Polynomial of a matrix, the exponential of a matrix, the functional calculus: an overview.

Cramer's rule for linear equation systems. Non-trivial solution for homogeneous systems. Eigenvalues and eigenvectors.

The characteristic polynomial of a matrix.

- Find eigenvalues and eigenvectors of the matrix

$$B = \begin{pmatrix} 3 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- If H is symmetric all the eigenvalues are real: prove this theorem in the 2×2 case.

Vector subspaces and their intersections.

- Intersection of eigenspaces w.r.t. different eigenvalues is $\{0\}$.
- Kronecker symbol. Orthonormal basis. Examples ....
- Orthogonal matrices preserve angles and lengths.
- Examples of orthogonal matrices: rotations in $\mathbb{R}^2$
- Suppose that the matrix A is symmetric and that λ, μ are distinct eigenvalues. If v is an eigenvector w.r.t. to λ and w is an eigenvector w.r.t. to μ then v, w are orthogonal. True or false?
- Diagonal matrices.
- Similar matrices.
- From orthonormal basis to orthogonal matrices.
- Diagonalizable matrices.
- The spectral theorem (no proof!): using orthonormal basis of eigenvectors it is possible to diagonalize symmetric matrices.
- Functional calculus for symmetric matrices.
- The square root of a symmetric positive semidefinite matrix.

Lecture 20 – Monday, October 9, 2023 (14:00-16:15)

Exercise Find the spectral decomposition of

$$A = \begin{pmatrix} \frac{5}{2} & \frac{3}{2} \\ \\ \frac{3}{2} & \frac{5}{2} \end{pmatrix}$$

and calculate $\sqrt{A}$.

Answer

$$\sqrt{A} = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \\ \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

- Definition of projection ($P^2 = P = P^t$)

- Examples of projections

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- Eigenvalues of a projection.

Open sets in the plane. Closed, bounded, compact sets. Weierstrass Theorem.

Lecture 21 – Tuesday, October 10, 2023 (14:00-16:15)

Partial derivatives.

Stationary points: how to study their nature using the eigenvalues of the Hessian matrix.

Exercise: study the stationary points of the function

$$f(x, y, z) = x^2 + y^4 + y^2 + z^3 - 2xz$$

Answer: (0,0,0) is a saddle point, $(\frac{2}{3}, 0, \frac{2}{3})$ is a local minimum

Directional derivatives.

Exercise. Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \left(\frac{x^2 y}{x^4 + y^2}\right)^2, & \text{if } (x, y) \neq (0, 0) \\ 0, & \text{if } (x, y) = (0, 0) \end{cases}$$

a) is not continuous at the origin;

b) has the partial derivatives at the origin;

c) has all the directional derivatives at the origin.

Lecture 22 – Wednesday, October 11, 2023 (14:00-16:15)

Characterization of linear transformations $T : \mathbb{R}^n \to \mathbb{R}$ as scalar products.

The notion of differentiable function and the tangent plane.

The gradient.

Stationary points.

If a function f is differentiable then f has directional derivatives in any direction and moreover

$$\langle \nabla f, v \rangle = D_v f$$

Continuity of a differentiable function by the Cauchy-Schwartz inequality. (No proof yet.)

Introduction to differential equations. The Cauchy problem. Linear differential equation of first order (homogeneous case): for an interval $I \subset \mathbb{R}$, $x_0 \in I$, $y_0 \in \mathbb{R}$ and a function $a(\cdot) \in \mathcal{C}(I)$ the unique solution of the Cauchy problem

$$\begin{cases} y' = ay & \text{in } I \\ y(x_0) = y_0 \end{cases}$$

is given by

$$y(x) = y_0 \exp\left(\int_{x_0}^{x} a(s)\mathrm{d}s\right).$$

If $X \sim N(0,1)$ its characteristic function $\varphi_X(t)$ satisfies the following Cauchy problem:

$$\begin{cases} y' = -xy & \text{in } \mathbb{R} \\ y(0) = 1 \end{cases}$$

From the above we deduce that $\varphi_X(t) = e^{-\frac{t^2}{2}}$.

Convex sets.

Brouwer's fixed point theorem: if $\emptyset \neq A \subset \mathbb{R}^n$ is compact, convex and $f : A \to A$ is continuous then there exist $x \in A$ such that $f(x) = x$. The idea of the proof in the general case.

Exemple: prove the Brouwer's fixed point theorem in the case $n = 1$ and $A = [0, 1]$.

Lecture 23 – Thursday, October 12, 2023 (14:00-16:15)

- Eigenspaces of a projection: the range $\mathrm{Range}(P)$ and the kernel $\mathrm{Ker}(P)$.
- $\mathrm{Range}(P) \perp \mathrm{Ker}(P)$
- Triangular matrices and linear system (forward and backward substitution).
- Determinant of a triangular matrix.
- If L is nonsingular and lower triangular then $A = LL^t$ is a symmetric positive definite matrix.

- Conversely (Cholesky decomposition): if A is symmetric positive definite there exists L non-singular lower triangular such that $A = LL^t$.
- Solving linear systems via Cholesky decomposition.
- Positive definite matrix: if A is a symmetric matrices then

$$\langle Av, v\rangle > 0 \quad \forall v \neq 0 \quad \longleftrightarrow \quad \text{eigenvalues of } A \text{ are positive}$$

- The geometric meaning of the determinant.
- If U is orthogonal then $\det(U) = \pm 1$. Viceversa not true. A counterexample:

$$A = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

We have that $A = A^t$ so that $AA^t = A^2 \neq I$.

Lecture 24 – Friday, October 13, 2023 (14:00-16:15)
The Jacobian matrix and its properties.

Regular curves. Two examples of non-regular curves

$$t \in [-1, 1] \qquad \gamma_1(t) = (t, |t|) \qquad \gamma_2(t) = (t^2, t^3)$$

The gradient of a function and its relation with the increasing-decreasing of the function.
Level curves.
Orthogonality of the gradient to level curves.

Lecture 25 – Monday, October 16, 2023 (14:00-16:00)

Constrained optimization. Lagrangian function and Lagrange multipliers.

- **Exercise.** Maximize (minimize) the function

$$f(x, y) = x^2 + y^2$$

subject to the constraint

$$g(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$$

using: a) parametrization of the curve; b) Lagrange multipliers.

The sufficient condition for differentiability in a point: existence of partial derivatives in a neighborhood and their continuity in the point.

Exercise. Show that the function

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

has the derivative everywhere in $\mathbb{R}$ but the derivative it is not continuous in 0.

Lecture 26 – Tuesday, October 17, 2023 (14:00-16:00)

Hyperplanes in $\mathbb{R}^n$: let $0 \neq p \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$. The *hyperplane generated by p and c* is the set $H_{p,c} := \{v \in \mathbb{R}^n | \langle p, v \rangle = c\}$.

The *half-space above* $H_{p,c}$ is the set $H^+_{p,c} := \{v \in \mathbb{R}^n | \langle p, v \rangle \geq c\}$.

The *half-space below* $H_{p,c}$ is the set $H^-_{p,c} := \{v \in \mathbb{R}^n | \langle p, v \rangle \leq c\}$.

An hyperplane $H_{p,c}$ separates two sets A, B if $A \subset H^+_{p,c}$ while $B \subset H^-_{p,c}$ (A and B are *on different sides* with respect to $H_{p,c}$).

The Separating Hyperplane Theorem: let $A, B \subset \mathbb{R}^n$ be two convex, disjoint sets. Then there exist a separating hyperplane for A and B.

- Diffeomorphisms in $\mathbb{R}$ and $\mathbb{R}^2$. The Jacobian matrix and its properties. The change of variable formula.

$$\int \int_{\mathbb{A}} g(x,y) \mathrm{d}x\mathrm{d}y = \int \int_{h^{-1}(A)} g(h(u,v)) |\det J_h(u,v)| \mathrm{d}u\mathrm{d}v$$

- Polar coordinates.

- Change of coordinates. The polar coordinates case:

$$\int \int_{\mathbb{R}^2} g(x,y) \mathrm{d}x\mathrm{d}y = \int_0^{2\pi} \int_0^{+\infty} f(\rho \cos\theta, \rho \sin\theta) \rho \, \mathrm{d}\rho\mathrm{d}\theta$$

- The formula

$$\int_{\mathbb{R}} e^{-x^2} \mathrm{d}x = \left(\int \int_{\mathbb{R}^2} e^{-(x^2+y^2)} \mathrm{d}x\mathrm{d}y \right)^{\frac{1}{2}} = \sqrt{\pi}$$

Lecture 27 – Thursday, October 19, 2023 (14:00-16:15)

The Taylor polynomial in two variables. If the Hessian is positive definite in a point then the point is a local minimum

The sufficient condition for the symmetry of the Hessian in a point: existence of mixed partial derivatives in a neighborhood and their continuity in the point (Schwartz-Young theorem).

Similar matrices have the same determinants. The determinant and the product of the eigenvalues. The trace and the sum of the eigenvalues.

For a function of two variable (under suitable differentiability conditions) in a stationary points where the Hessian determinant is positive the condition $f_{xx} > 0$ implies that the point is a local minimum (and a local maximum if $f_{xx} < 0$).

Quasiconvex Functions

Proposition 0.1. *Let U be an interval (in $\mathbb{R}$) and $f : U \to \mathbb{R}$.*
The following properties are equivalent
1.

$$f(x) \leq f(y) \Rightarrow f(tx + (1-t)y) \leq f(y) \qquad x, y \in U, \quad t \in [0,1].$$

2.

$$f(tx + (1-t)y) \leq \text{Max}\{f(x), f(y)\} \qquad x, y \in U, \quad t \in [0,1].$$

3.
The set $C_a^- := \{x \in U | f(x) \leq a\}$ is convex for all $a \in \mathbb{R}$.

In such a case we say that the function f is *quasiconvex*.
Proof
1. $\to$ 2.
2. $\to$ 3.
3. $\to$ 1.

Proposition 0.2. *f convex implies f quasiconvex.*

Proposition 0.3. *f non decreasing implies f quasiconvex.*

Therefore $\log(x)$ (which is concave) is quasiconvex.

Proposition 0.4. *If f decreases monotonically until it reaches a global minimum and then monotonically rises then f quasiconvex.*

Therefore $-e^{-x^2}$ (which is neither concave nor convex) is quasiconvex.
Exercise 1.
(X, Y) is a random vector with uniform density on $A = \{(x, y) \in R^2 | x^2 + y^2 \leq 1\}$.
Let $U := X + Y$.
ii) Calculate $F_U(-2)$, $F_U(0)$, $F_U(2)$.
ii) Find the marginal densities f_X, f_Y.
i) Are X and Y independent?

Lecture 28 – Monday, October 23, 2023 (14:00-16:15)

Introduction to Kuhn-Tucker Theorem

- Optimization w.r.t. inequality constraints.
- Binding constraints.
- Complementary slackness condition.
- Examples of the Simon-Blume: 18.7 at page 428 (using the Lagrangian formulation).
- Examples of the Simon-Blume: 18.7 at page 428 (using the parameterization of the boundary).
- Examples of the Simon-Blume:18.9 at page 431.

Exercise. Maximize $f(x,y) = x^2 - 2x + y^2$ subject to the constraint $\frac{x^2}{4} + y^2 \leq 1$ using the Lagrangian and using the parametrization of the boundary of the constraint set. (Solution. You find with both methods five candidates: $(1,0), (2,0)(-2,0)(\frac{4}{3}, \frac{\sqrt{5}}{3})(\frac{4}{3}, -\frac{\sqrt{5}}{3})$. The global maximum is in the point (-2,0). To better understand the solution you may write the function as $f(x,y) = (x-1)^2 + y^2 - 1$.)

Lecture 29 – Tuesday, October 24, 2023 (14:00-16:15)

- Sets in the plane that are cartesian products.
- The support of a function.
- A geometric condition for the independence of marginals.

Exercise. Suppose that P is a projection.
1. Prove that $\tilde{P} := I - P$ is a projection.
2. Prove that Ker(P)=Range($\tilde{P}$).
3. Prove that Range(P)=Ker($\tilde{P}$).
4. $P\tilde{P} = \tilde{P}P = 0$.

- Examples of the Simon-Blume: 18.10 at page 435 (using the Lagrangian formulation).
- Examples of the Simon-Blume: 18.10 at page 435 (using the parameterization of the boundary).

Lecture 30 – Wednesday, October 25, 2023 (14:00-16:15)

- If 0 is an eigenvalue of a linear transformation T then T is not injective. True or false?
- Every matrix $A \neq 0$ has an inverse. True or false?
- If $X, Y \sim N(0, \sigma^2)$ then $\frac{X}{\sqrt{X^2+Y^2}}$ does not depend on σ.
- The characteristic function of a random variable with symmetric distribution is real.
- If $(X, Y) \sim \mathcal{N}(b, \Gamma)$ and

$$\begin{pmatrix} U \\ V \end{pmatrix} = B \begin{pmatrix} X \\ Y \end{pmatrix} + c$$

then $(U, V) \sim \mathcal{N}(Bb + c, B\Gamma B^t)$.

Exercise. Prove that the matrix

$$C = \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix}$$

is positive definite. Then find the Choleski decompostion of C. Answer:

$$C = \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}.$$

- Transpose, cofactor, adjugate, determinant, inverse matrix and their properties for 2×2 matrices.
- The Jacobian of an affine transformation.
- How the density changes under a transformation (g diffeomorphism) of a random vector.

$$f_{g(X,Y)}(x, y) = f_{X,Y}(g^{-1}(x, y))|\det J_{g^{-1}}(x, y)|$$

- Example:

$$f_{aX+b}(x) = \frac{1}{|a|} \cdot f_X\left(\frac{x-b}{a}\right)$$

- The density of a bivariate gaussian (The proof!):

$$f_{X_1,X_2}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$
$$\times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\cdot\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

Lectures 31 and 32 – Thursday, October 26, 2023 (11:00-13:00 &14:00-16:15)

Solution of Simulations 1, 2, 3, 4.

- $X \sim B(1,p)$, $Y \sim B(1,p)$ and $X \perp Y$ implies $\mathbb{E}(X|X+Y) = \frac{1}{2}(X+Y)$. Once you prove the result check the law of iterated expectation.
- $X \sim$ Poisson(λ), $Y \sim$ Poisson(μ) and $X \perp Y$ implies $\mathbb{E}(X|X+Y) = \frac{\lambda}{\lambda+\mu}(X+Y)$. Once you prove the result check the law of iterated expectation.
- Remark: $E(X|Y)$ depends only on the joint distribution of X and Y (besides the distributions of X, Y).
- $X_1, X_2, ..., X_n$ i.i.d.r.v.s and $S_n = X_1 + X_2 + \cdots X_n$. Then $E(X_1|S_n) = \frac{S_n}{n}$.

Lecture 33 – Friday, October 27, 2023 (11:00-13:00)

Convex sets in $\mathbb{R}$ are intervals or half lines.

Quasiconvex functions on $\mathbb{R}^n$.

Proposition. Let U be an open convex set in $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ a $\mathcal{C}^1$ function; then f is quasiconvex if and only if

$$f(y) \leq f(x) \quad \Rightarrow \quad \langle \nabla f(x), y - x\rangle \leq 0.$$

The example of the logarithm.

Quasilinear functions.

Maximal rank for a matrix

Non-degenerate constraint qualification (NDCQ): the Jacobian of the binding constraint has maximal rank.

Check the NDCQ for the exercises done on the Simon-Blume.

A counterexample (without NDCQ the Lagrangian formulation does not work). Consider the following Problem. the function

$$f(x, y) = -x$$

subject to the constraints

$$g_1(x, y) = -(x-1)^3 - y^2 \geq 0$$

$$g_2(x, y) = xy \geq 0$$

Solution. The function f has a global minimum in

$$p^* = (x^*, y^*) = (1, 0).$$

In p^* both the constraints are binding and their Jacobian is

$$\begin{pmatrix} \nabla g_1(p^*) \\ \nabla g_2(p^*) \end{pmatrix} = \begin{pmatrix} -3(x^*-1)^2 & -2y^* \\ y^* & x^* \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

which has not maximal rank. Therefore the NDCQ (non degenerate constraints qualification) does not hold. Indeed for the Lagrangian

$$\mathcal{L}(p, \lambda) = \mathcal{L}(x, y, \lambda_1, \lambda_2) = -x - \lambda_1(-(x-1)^3 - y^2) - \lambda_2 xy$$

one has

$$\frac{\partial}{\partial x}\mathcal{L}(p^*, \lambda) = -1 - \lambda_1(-3(1-1)^2) - \lambda_2 \cdot 0 = -1 \neq 0 \quad \text{for} \quad \text{any} \quad \lambda = (\lambda_1, \lambda_2).$$