Complete Mathematics

Miko-pedia AI

May 4, 2026

Contents

Overview. This complete note provides a unified undergraduate review of core mathematical topics, from logic and discrete structures through calculus, linear algebra, probability, and modern analysis. Each section is designed to stand alone for exam review: definitions, key theorems, representative examples, and concise summaries are included throughout. The sections follow a natural progression from foundational reasoning to advanced topics such as topology and numerical methods. Use this overview to locate weak areas and target your study efficiently.

1 The Language of Mathematics: Propositions, Sets, Functions, and Proofs

1.1 Core ideas

Mathematics rests on a foundation of logic and set theory. A **proposition** is a declarative statement that is either true or false (e.g., "2+2 = 4" is true; "2+2 = 5" is false). Propositions are combined using logical connectives: $\neg$ (negation, "not"), $\wedge$ (conjunction, "and"), $\vee$ (disjunction, "or"), $\implies$ (implication, "if...then"), and $\iff$ (biconditional, "if and only if"). An implication $P \implies Q$ is false only when P is true and Q is false; it is vacuously true when P is false. A **predicate** $P(x)$ is a statement whose truth depends on the variable x, quantified by $\forall$ ("for all") or $\exists$ ("there exists"). De Morgan's laws for quantifiers state $\neg(\forall x\, P(x)) \iff \exists x\, \neg P(x)$ and $\neg(\exists x\, P(x)) \iff \forall x\, \neg P(x)$.

A **set** is a collection of distinct objects, written $\{a, b, c, \dots\}$. The empty set $\varnothing$ contains no elements. Basic operations include union $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$, intersection $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$, and complement $A^c = \{x \mid x \notin A\}$ relative to a universe U. The Cartesian product $A \times B = \{(a, b) \mid a \in A,\ b \in B\}$ produces ordered pairs. A relation on A and B is a subset of $A \times B$.

A **function** $f : A \to B$ assigns each $a \in A$ exactly one $b \in B$, written $f(a) = b$. A is the domain, B the codomain, and $\{f(a) \mid a \in A\}$ the range/image. A function is **injective** (one-to-one) if $f(a_1) = f(a_2)$ implies $a_1 = a_2$; **surjective** (onto) if every $b \in B$ has some $a \in A$ with $f(a) = b$; and **bijective** if both hold. Bijections have inverses $f^{-1} : B \to A$.

Proof techniques are the tools of mathematical reasoning. **Direct proof**: assume hypotheses, deduce conclusion via logical steps. **Proof by contrapositive**: prove $P \implies Q$ by showing $\neg Q \implies \neg P$. **Proof by contradiction**: assume the negation of what you want to prove and derive a contradiction. **Mathematical induction**: prove $P(1)$ (base case) and $P(k) \implies P(k+1)$ (inductive step) for all $n \in \mathbb{N}$.

1. **Proposition**: statement with truth value; connectives $\neg, \wedge, \vee, \implies, \iff$.
2. **Set**: unordered collection; operations $\cup, \cap, \setminus, \times$.
3. **Function**: rule $f : A \to B$; types: injective, surjective, bijective.
4. **Proof methods**: direct, contrapositive, contradiction, induction.

For review, be able to: translate English statements into logical form; build truth tables; negate compound statements; use quantifiers correctly; perform set operations; prove function injectivity/surjectivity; choose and execute the appropriate proof technique. Identify the logical structure, the quantifier order, the inductive hypothesis, and the contradiction being derived.

1.2 Mathematical spine

Truth tables:

P	Q	$P \wedge Q$	$P \implies Q$
T	T	T	T
T	F	F	F
F	T	F	T
F	F	F	T

$$\text{De Morgan: } \neg(P \wedge Q) \iff \neg P \vee \neg Q, \qquad \neg(P \vee Q) \iff \neg P \wedge \neg Q$$

$$\text{Induction: } P(1) \wedge [\forall k\, (P(k) \implies P(k+1))] \implies \forall n\, P(n)$$

Section summary The language of mathematics comprises propositions (with connectives and quantifiers), sets (with union, intersection, Cartesian product), and functions (injective, surjective, bijective). Proof techniques — direct, contrapositive, contradiction, induction — form the backbone of mathematical argumentation. Mastery of this language is prerequisite for all higher mathematics.

2 Discrete Structures: Induction, Combinatorics, Graphs, and Algorithms

2.1 Core ideas

Discrete mathematics studies countable, separated structures. **Mathematical induction** (strong and weak) proves statements for all natural numbers. **Recursion** defines objects in terms of themselves, e.g., the Fibonacci sequence $F_1 = 1$, $F_2 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n \geq 3$.

Combinatorics counts arrangements and selections. The **multiplication principle**: if task A has m ways and task B has n ways, the pair (A, B) has $m \cdot n$ ways. A **permutation** is an ordered arrangement: $P(n, k) = n!/(n-k)!$ ways to choose k items from n in order. A **combination** is an unordered selection: $\binom{n}{k} = n!/(k!(n-k)!)$. The **Binomial Theorem** states $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$. The **Pigeonhole Principle**: if n items go into m boxes and $n > m$, at least one box contains ≥ 2 items. More generally, if $n > m \cdot k$, some box contains $\geq k+1$ items.

A **graph** $G = (V, E)$ consists of vertices V and edges E (unordered pairs of vertices). A graph is **simple** if no loops or multiple edges. The **degree** $\deg(v)$ is the number of edges incident to v. The Handshaking Lemma: $\sum_{v \in V} \deg(v) = 2|E|$. A **path** is a sequence of distinct vertices each adjacent to the next; a **cycle** is a closed path. A graph is **connected** if any two vertices have a path between them. A **tree** is a connected acyclic graph; any tree on n vertices has exactly $n-1$ edges. An **Eulerian circuit** traverses every edge exactly once and exists iff every vertex has even degree. A **Hamiltonian cycle** visits every vertex exactly once (no simple characterization). Graph coloring assigns colors to vertices so adjacent vertices differ; the **chromatic number** $\chi(G)$ is the minimum colors needed.

Algorithms are step-by-step procedures. The **Euclidean algorithm** finds $\gcd(a, b)$ by repeated division. Big-O notation describes asymptotic growth: $f(n) = O(g(n))$ if $|f(n)| \leq c|g(n)|$ for $n \geq n_0$. For example, searching a sorted list via binary search is $O(\log n)$, while bubble sort is $O(n^2)$.

Graph algorithms. **Breadth-first search** (BFS) explores a graph level by level from a source vertex, using a queue; it finds shortest paths in unweighted graphs. **Depth-first search** (DFS) explores as far as possible along each branch before backtracking, using a stack (or recursion); it is used for topological sorting and detecting cycles. **Dijkstra's algorithm** finds shortest paths from a source to all other vertices in a weighted graph with non-negative edge weights by maintaining a priority queue of tentative distances and greedily selecting the minimum.

Strong induction (complete induction) assumes the statement holds for all integers up to k to prove it for $k+1$. It is useful when the truth of $P(k+1)$ depends on multiple preceding cases. For example, to prove that every integer $n \geq 2$ can be written as a product of primes: base case $n = 2$ is prime. Assume all integers from 2 to k factor into primes. For $k+1$, either it is prime (done) or it factors as ab with $2 \leq a, b \leq k$, and the induction hypothesis applies to a and b.

1. **Induction**: prove base case, assume for k, prove for $k+1$; strong induction assumes all cases $\leq k$.

2. **Permutations** $P(n,k) = n!/(n-k)!$; **combinations** $\binom{n}{k}$.

3. **Graph**: $G = (V, E)$; Eulerian circuits (even degrees), trees ($n-1$ edges), BFS/DFS traversal.

4. **Algorithms**: Euclidean, Dijkstra; Big-O measures worst-case complexity.

For review, be able to: set up and solve recurrence relations; compute permutations and combinations; apply the Binomial Theorem; prove graph properties using degree sums; determine Eulerian/Hamiltonian existence; analyze algorithm runtime using Big-O; trace BFS and DFS on a small graph; apply strong induction. Identify the recursive structure, the counting principle, the parity argument, and the dominant term in complexity.

2.2 Mathematical spine

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \qquad (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k, \qquad \sum_{v \in V} \deg(v) = 2|E|$$

Section summary Discrete structures encompass induction (for proving $\mathbb{N}$-indexed statements), combinatorics (permutations $\binom{n}{k}$, combinations $P(n,k)$, Binomial Theorem), graph theory (vertices, edges, Eulerian/Hamiltonian paths, trees), and algorithm analysis (Big-O, Euclidean algorithm). These tools are essential for computer science, optimization, and reasoning about finite structures.

3 Single Variable Calculus: Change, Approximation, and Integration

3.1 Core ideas

Calculus studies continuous change. The **limit** $\lim_{x \to a} f(x) = L$ means $f(x)$ can be made arbitrarily close to L by taking x sufficiently close to a (but not equal). **One-sided limits** consider approach from left ($x \to a^-$) or right ($x \to a^+$). A function is **continuous** at a if $\lim_{x \to a} f(x) = f(a)$. The **Intermediate Value Theorem**: if f is continuous on $[a, b]$ and y is between $f(a)$ and $f(b)$, then $f(c) = y$ for some $c \in (a, b)$.

The **derivative** measures instantaneous rate of change:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Geometrically, it is the slope of the tangent line at $x = a$. Differentiability implies continuity. Key differentiation rules: **power rule** $\frac{d}{dx} x^n = n x^{n-1}$; **product rule** $(fg)' = f'g + fg'$; **quotient rule** $(f/g)' = (f'g - fg')/g^2$; **chain rule** $(f \circ g)'(x) = f'(g(x))\, g'(x)$. Derivatives of elementary functions: $\frac{d}{dx} \sin x = \cos x$, $\frac{d}{dx} \cos x = -\sin x$, $\frac{d}{dx} e^x = e^x$, $\frac{d}{dx} \ln x = 1/x$. The **Mean Value Theorem**: if f is continuous on $[a, b]$ and differentiable on (a, b), then $f'(c) = (f(b) - f(a))/(b-a)$ for some $c \in (a, b)$. Derivatives enable **optimization** (finding extrema via critical points), **linear approximation** $f(x) \approx f(a) + f'(a)(x - a)$, and **L'Hôpital's rule** for limits of indeterminate forms.

Integration reverses differentiation. The **indefinite integral** $\int f(x)\,dx = F(x) + C$ means $F'(x) = f(x)$. The **definite integral** $\int_a^b f(x)\,dx$ gives the signed area under the curve. The **Fundamental Theorem of Calculus** links the two: Part 1 says $\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$; Part 2 says $\int_a^b f(x)\,dx = F(b) - F(a)$ where $F' = f$. Integration techniques include **substitution** ($u = g(x)$), **integration by parts** ($\int u\,dv = uv - \int v\,du$), and **partial fractions**. Applications include area between curves, volume of revolution (disks, shells), arc length, and work.

1. **Limit**: $\lim_{x \to a} f(x) = L$; continuity, IVT.
2. **Derivative**: $f'(a) = \lim_{h \to 0}(f(a + h) - f(a))/h$; product, quotient, chain rules.
3. **Integral**: $\int_a^b f(x)\,dx$; FTC; substitution, integration by parts.
4. **Applications**: optimization, linear approximation, area, volume.

For review, be able to: compute limits using algebra and L'Hôpital; differentiate any elementary function; apply chain rule in multiple compositions; solve optimization problems; compute definite and indefinite integrals using standard techniques; apply FTC; set up integrals for area and volume. Identify the indeterminate form, the composition structure, the critical points, and the region of integration.

3.2 Mathematical spine

$$\boxed{f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}}, \qquad \boxed{\int_a^b f(x)\,dx = F(b) - F(a)}, \qquad \frac{d}{dx}\int_a^x f(t)\,dt = f(x)$$

Section summary Single-variable calculus provides the language for describing change. Limits formalize approximation; derivatives measure instantaneous rates and enable optimization; integrals accumulate quantities. The Fundamental Theorem of Calculus unifies these ideas, forming the core of mathematical analysis and its applications in science and engineering.

4 Linear Algebra: Spaces, Maps, Eigenvalues, and Projections

4.1 Core ideas

Linear algebra studies vector spaces and linear transformations. A **vector space** V over a field $\mathbb{F}$ (typically $\mathbb{R}$) is a set closed under addition and scalar multiplication, satisfying associativity, commutativity, distributivity, existence of zero vector $\mathbf{0}$, and additive inverses. A **subspace** $U \subseteq V$ is a nonempty subset closed under the same operations. The **span** of vectors $\{v_1, \ldots, v_k\}$ is the set of all linear combinations $\sum c_i v_i$. A set is **linearly independent** if $\sum c_i v_i = \mathbf{0}$ implies all $c_i = 0$. A **basis** is a linearly independent spanning set; the number of basis vectors is the **dimension** $\dim V$.

A **linear transformation** $T : V \to W$ satisfies $T(u + v) = T(u) + T(v)$ and $T(cv) = cT(v)$. The **kernel** (null space) $\ker T = \{v \mid T(v) = 0\}$ and **image** (range) $\operatorname{im} T = \{T(v) \mid v \in V\}$ are subspaces. The **rank-nullity theorem**: $\dim \ker T + \dim \operatorname{im} T = \dim V$.

A matrix $A \in \mathbb{R}^{m \times n}$ represents a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ via $x \mapsto Ax$. Matrix operations: addition (componentwise), multiplication $(AB)_{ik} = \sum_j A_{ij} B_{jk}$, and transpose $(A^T)_{ij} = A_{ji}$. The **determinant** $\det A$ for a square matrix measures volume scaling and invertibility: A is invertible iff $\det A \neq 0$.

An **eigenvector** of A satisfies $Av = \lambda v$ for scalar λ (the **eigenvalue**). Eigenvalues solve $\det(A - \lambda I) = 0$. A matrix is **diagonalizable** if it has n linearly independent eigenvectors. The **spectral theorem**: every real symmetric matrix has an orthonormal basis of eigenvectors with real eigenvalues. The **singular value decomposition** (SVD) factorizes any $A \in \mathbb{R}^{m\times n}$ as $A = U\Sigma V^T$ where U, V are orthogonal and Σ is diagonal with singular values $\sigma_i \geq 0$.

Inner products generalize dot products: $\langle x, y\rangle = x^T y$ in $\mathbb{R}^n$. Orthogonal projection onto a subspace U with orthonormal basis $\{u_1, \ldots, u_k\}$ is $P_U(v) = \sum \langle v, u_i\rangle u_i$. The **Gram-Schmidt process** produces an orthonormal basis from any basis. The **least squares** solution to $Ax = b$ minimizes $\|Ax - b\|^2$ via $A^T Ax = A^T b$.

1. **Vector space**: basis, dimension, subspace, span, linear independence.
2. **Linear map**: kernel, image, rank-nullity; matrix representation.
3. **Eigenvalues**: $\det(A - \lambda I) = 0$; diagonalization; SVD.
4. **Projections**: orthogonal projection, Gram-Schmidt, least squares.

For review, be able to: check linear independence; find basis and dimension; compute matrix products and determinants; solve eigenvalue problems; diagonalize symmetric matrices; compute SVD; project vectors onto subspaces; solve least squares problems. Identify the vector space structure, the matrix factorization, the eigenspace geometry, and the orthogonality relations.

4.2 Mathematical spine

$$\boxed{Av = \lambda v}, \quad \det(A - \lambda I) = 0, \quad A = U\Sigma V^T, \quad \mathrm{proj}_U(v) = \sum_{i=1}^{k} \langle v, u_i\rangle u_i$$

Section summary Linear algebra unifies vector spaces and their transformations. Key concepts: bases and dimension (coordinates); linear maps and matrices (representations); eigenvalues and eigenvectors (diagonalization, SVD); inner products and projections (orthogonality, least squares). This framework underlies nearly all applied mathematics, from differential equations to machine learning.

5 Multivariable Analysis: Gradients, Multiple Integrals, and Vector Fields

5.1 Core ideas

Multivariable calculus extends single-variable ideas to functions of several variables. For $f : \mathbb{R}^n \to \mathbb{R}$, the **partial derivative** $\partial f/\partial x_i$ at a is the derivative along the x_i direction. The **gradient** $\nabla f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_n)$ points in the direction of steepest ascent. The **directional derivative** $D_u f = \nabla f \cdot u$ gives the rate of change in direction u.

For $f : \mathbb{R}^n \to \mathbb{R}^m$, the **Jacobian matrix** J has entries $J_{ij} = \partial f_i/\partial x_j$. The **chain rule**: $D(f \circ g)(x) = Df(g(x))\, Dg(x)$. **Taylor's theorem** in several variables: $f(x + h) = f(x) + \nabla f(x) \cdot h + \frac{1}{2} h^T H_f(x) h + \cdots$ where H_f is the Hessian matrix $H_{ij} = \partial^2 f/\partial x_i \partial x_j$. Critical points satisfy $\nabla f = 0$; their nature (min/max/saddle) is determined by the eigenvalues of H_f (second derivative test). **Lagrange multipliers** find extrema subject to constraints $g(x) = 0$: solve $\nabla f = \lambda \nabla g$ and $g(x) = 0$.

Multiple integrals extend integration to higher dimensions. The double integral $\iint_R f(x, y)\, dA$ is the volume under $z = f(x, y)$ over region R. Iterated integrals: $\iint_R f\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\, dx$. Change of variables: $\iint_R f(x, y)\, dx\, dy = \iint_S f(u, v)\, |\det J|\, du\, dv$; polar coordinates (r, θ) give

$dA = r\,dr\,d\theta$. Triple integrals in cylindrical ($dV = r\,dr\,d\theta\,dz$) and spherical coordinates ($dV = \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta$) handle 3D volumes.

A **vector field** $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ assigns a vector to each point. The **divergence** $\nabla \cdot \mathbf{F} = \partial P/\partial x + \partial Q/\partial y + \partial R/\partial z$ measures outflow. The **curl** $\nabla \times \mathbf{F}$ measures rotation. A line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$ accumulates the tangential component along curve C. For conservative fields $\mathbf{F} = \nabla\phi$, the line integral is path-independent and $\int_C \nabla\phi \cdot d\mathbf{r} = \phi(B) - \phi(A)$ (Fundamental Theorem for Line Integrals). The big integral theorems: **Green's Theorem** $\oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_D (\nabla \times \mathbf{F}) \cdot \mathbf{k}\,dA$ (2D), **Stokes' Theorem** $\oint_{\partial S} \mathbf{F} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S}$, and the **Divergence Theorem** $\iint_{\partial V} \mathbf{F} \cdot d\mathbf{S} = \iiint_V (\nabla \cdot \mathbf{F})\,dV$.

1. **Gradient**: ∇f; directional derivative; Hessian; Lagrange multipliers.
2. **Integration**: double/triple integrals; change of variables (polar, cylindrical, spherical).
3. **Vector fields**: divergence, curl, line integrals, conservative fields.
4. **Integral theorems**: Green, Stokes, Divergence.

Example (Green's theorem). Let C be the unit circle $x^2 + y^2 = 1$ traversed counter-clockwise and let $\mathbf{F} = (-y, x)$. Then $\nabla \times \mathbf{F} = \partial_x(x) - \partial_y(-y) = 2$. By Green's theorem,

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_D 2\,dA = 2\pi,$$

since the area of the unit disk is π. This is much simpler than parameterizing the circle and computing the line integral directly.

For review, be able to: compute gradients and directional derivatives; find and classify critical points; set up and evaluate multiple integrals in various coordinate systems; compute line and surface integrals; apply Green/Stokes/Divergence theorems. Identify the constraint surface, the coordinate system simplifying the region, the conservative potential, and the boundary orientation.

5.2 Mathematical spine

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right), \quad \iint_R f\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f\,dy\,dx, \quad \oint_{\partial S} \mathbf{F} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S}$$

Section summary Multivariable analysis extends calculus to higher dimensions. Gradients and Hessians characterize local behavior; multiple integrals compute volumes and averages; vector field calculus culminates in Green's, Stokes', and the Divergence theorems, which relate local derivatives to global boundary integrals. These tools are essential for physics (electromagnetism, fluid dynamics) and optimization.

6 Probability: Conditional Probability, Random Variables, and Distributions

6.1 Core ideas

Probability quantifies uncertainty. A **sample space** Ω is the set of all outcomes. An **event** $A \subseteq \Omega$ is a subset of outcomes. A **probability measure** Pr satisfies: $\Pr(\Omega) = 1$, $0 \le \Pr(A) \le 1$, and $\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$ for disjoint events. The **complement** rule: $\Pr(A^c) = 1 - \Pr(A)$. The **inclusion-exclusion principle**: $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

Conditional probability $\Pr(A \mid B) = \Pr(A \cap B)/\Pr(B)$ updates probability given that B occurred. Events A and B are **independent** if $\Pr(A \cap B) = \Pr(A)\Pr(B)$, equivalently

$\Pr(A \mid B) = \Pr(A)$. **Bayes' theorem**: $\Pr(A \mid B) = \Pr(B \mid A)\Pr(A)/\Pr(B)$, which reverses the conditioning. The **law of total probability**: $\Pr(B) = \sum_i \Pr(B \mid A_i)\Pr(A_i)$ for a partition $\{A_i\}$.

A **random variable** $X : \Omega \to \mathbb{R}$ assigns numbers to outcomes. The **cumulative distribution function** (CDF) $F_X(x) = \Pr(X \le x)$. For discrete X, the **probability mass function** (PMF) $p_X(x) = \Pr(X = x)$ satisfies $\sum_x p_X(x) = 1$. For continuous X, the **probability density function** (PDF) $f_X(x)$ satisfies $\Pr(a \le X \le b) = \int_a^b f_X(x)\,dx$ and $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$. The **expected value** $E[X] = \sum xp(x)$ (discrete) or $\int xf(x)\,dx$ (continuous). The **variance** $\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - \mu^2$; $\sigma = \sqrt{\mathrm{Var}(X)}$ is the standard deviation.

Key discrete distributions: **Bernoulli**(p): $X \in \{0,1\}$, $\Pr(X=1) = p$. **Binomial**(n,p): sum of n i.i.d. Bernoulli, $p(k) = \binom{n}{k}p^k(1-p)^{n-k}$. **Poisson**$(\lambda)$: $p(k) = e^{-\lambda}\lambda^k/k!$, $k = 0,1,\ldots$. Key continuous distributions: **Uniform**(a,b): $f(x) = 1/(b-a)$ for $x \in [a,b]$. **Normal**(μ,σ^2): $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$. **Exponential**$(\lambda)$: $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$. The **central limit theorem** (CLT): the sum of n i.i.d. random variables with mean μ and variance σ^2 is approximately $\mathrm{Normal}(n\mu, n\sigma^2)$ for large n.

The **moment generating function** $\phi_X(t) = E[e^{tX}]$ uniquely determines the distribution. **Covariance** $\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$ measures joint linear dependence; correlation $\rho = \mathrm{Cov}(X,Y)/(\sigma_X\sigma_Y)$ lies in $[-1,1]$.

1. **Conditional probability**: $\Pr(A \mid B) = \Pr(A \cap B)/\Pr(B)$; Bayes' theorem.
2. **Random variables**: PMF/PDF, CDF, expectation, variance.
3. **Distributions**: Bernoulli, Binomial, Poisson, Normal, Exponential.
4. **CLT**: sums of i.i.d. variables converge in distribution to Normal.

For review, be able to: compute probabilities using Bayes' rule; calculate expectations and variances; recognize which distribution applies to a given experiment; apply CLT to approximate sums; compute and interpret covariance and correlation. Identify the independence structure, the distribution family, the normal approximation, and the moment generating function.

6.2 Mathematical spine

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \qquad E[X] = \int_{-\infty}^{\infty} xf(x)\,dx, \qquad \mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Section summary Probability provides the mathematical framework for uncertainty. Conditional probability and Bayes' theorem enable reasoning under partial information. Random variables and their distributions (Binomial, Poisson, Normal, etc.) model real-world randomness. Expectation, variance, and the Central Limit Theorem form the bridge from probability to statistics.

7 Statistics: Estimation, Testing, Regression, and Inference

7.1 Core ideas

Statistics uses data to make inferences about populations. A **parameter** is a numerical characteristic of a population; a **statistic** is a numerical characteristic of a sample. **Point estimation** produces a single best guess; **interval estimation** gives a plausible range. **Maximum likelihood estimation** (MLE) chooses θ maximizing the likelihood $L(\theta) = \prod_{i=1}^n f(x_i \mid \theta)$. For example, the MLE of the mean μ for Normal data is $\hat{\mu} = \bar{x} = \frac{1}{n}\sum x_i$.

The **sampling distribution** of a statistic is its probability distribution over repeated samples. The **standard error** is the standard deviation of the sampling distribution. For i.i.d. data with variance σ^2, $\mathrm{SE}(\bar{X}) = \sigma/\sqrt{n}$. A **confidence interval** for the mean when σ is known: $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$, where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal. When σ is unknown, use the t-distribution: $\bar{x} \pm t_{\alpha/2,\, n-1} \cdot s/\sqrt{n}$ where s is the sample standard deviation.

Hypothesis testing assesses evidence against a null hypothesis H_0. The p-value is Pr(observed or more ext H_0). If $p < \alpha$ (significance level), reject H_0. A **Type I error** rejects a true H_0; **Type II error** fails to reject a false H_0. The z-test for a mean: $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$. The t-test uses $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$.

Linear regression models $Y = \beta_0 + \beta_1 X + \varepsilon$, with $\varepsilon \sim N(0, \sigma^2)$ i.i.d. The least squares estimates minimize $\sum(y_i - \beta_0 - \beta_1 x_i)^2$: $\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$. The R^2 statistic measures the proportion of variance explained. Multiple regression extends to p predictors: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.

Bayesian inference treats parameters as random variables with prior $\pi(\theta)$ and posterior $\pi(\theta \mid \mathbf{x}) \propto f(\mathbf{x} \mid \theta)\,\pi(\theta)$. The posterior mean serves as a Bayesian estimator, and credible intervals are the Bayesian analog of confidence intervals.

1. **Estimation**: MLE, confidence intervals for mean (z and t).
2. **Testing**: null/alternative, p-value, z-test, t-test, Type I/II errors.
3. **Regression**: least squares; $\hat{\beta} = (X^TX)^{-1}X^Ty$; R^2.
4. **Bayesian**: posterior $\propto$ likelihood $\times$ prior.

For review, be able to: compute MLEs for common distributions; construct and interpret confidence intervals; perform z- and t-tests; interpret p-values; fit and interpret linear regression; derive posterior distributions for simple conjugate priors. Identify the estimator, the test statistic null distribution, the regression assumptions, and the prior-posterior relationship.

7.2 Mathematical spine

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Section summary Statistics bridges data and decisions through estimation, testing, and modeling. Point estimates (MLE), interval estimates (confidence intervals), and hypothesis tests (z, t) form classical inference. Linear regression models relationships. Bayesian methods incorporate prior knowledge via Bayes' theorem. These tools are the foundation of data analysis and scientific discovery.

8 Real Analysis: Limits, Continuity, Convergence, and Rigorous Integration

8.1 Core ideas

Real analysis provides rigorous foundations for calculus. The real numbers $\mathbb{R}$ form a **complete ordered field**: they are ordered (totally), form a field under $+$ and $\times$, and satisfy the **completeness axiom** (every nonempty set bounded above has a supremum). A sequence $\{a_n\}$ **converges** to L if $\forall\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $n \geq N \implies |a_n - L| < \varepsilon$. A sequence is **Cauchy** if $\forall\varepsilon > 0$, $\exists N$ such that $m, n \geq N \implies |a_m - a_n| < \varepsilon$. In $\mathbb{R}$, a sequence converges iff it is Cauchy. The **Bolzano-Weierstrass theorem**: every bounded sequence in $\mathbb{R}$ has a convergent subsequence.

For a function $f : \mathbb{R} \to \mathbb{R}$, $\lim_{x\to a} f(x) = L$ means $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon$. f is **continuous** at a if $\lim_{x\to a} f(x) = f(a)$. Equivalent: preimages of open sets are open. The **Intermediate Value Theorem**: a continuous function on $[a, b]$ attains every value between $f(a)$ and $f(b)$. The **Extreme Value Theorem**: a continuous function on a closed bounded interval attains its maximum and minimum.

Uniform continuity strengthens continuity: $\forall \varepsilon > 0$, $\exists \delta > 0$ that works for all x simultaneously (independent of the point). On closed bounded intervals, continuity implies uniform continuity (Heine-Cantor theorem). A sequence of functions $\{f_n\}$ **converges pointwise** to f if $f_n(x) \to f(x)$ for each x, and **uniformly** if $\sup_x |f_n(x) - f(x)| \to 0$. Uniform convergence preserves continuity and commutes with integration.

Riemann integration: the integral $\int_a^b f(x)\,dx$ is the limit of Riemann sums $\sum f(x_i^*)\Delta x_i$ as the partition mesh goes to zero. A function is Riemann integrable if the upper and lower sums converge (which happens if the set of discontinuities has measure zero). The **Fundamental Theorem of Calculus** (rigorous version): if f is integrable on $[a, b]$ and $F'(x) = f(x)$ uniformly, then $\int_a^b f = F(b) - F(a)$. The **Lebesgue integral** generalizes Riemann integration by partitioning the range rather than the domain, allowing integration of more functions and better convergence theorems (dominated convergence, monotone convergence).

Series $\sum_{n=1}^{\infty} a_n$ converge if the partial sums converge. Tests: ratio test, root test, integral test, comparison test. Power series $\sum c_n(x - a)^n$ have a radius of convergence $R = 1/\limsup \sqrt[n]{|c_n|}$. Within the radius, they can be differentiated and integrated termwise.

1. **Sequences**: ε-N definition; Cauchy criterion; Bolzano-Weierstrass.
2. **Functions**: ε-δ continuity; uniform continuity; IVT, EVT.
3. **Integration**: Riemann sums; FTC; Lebesgue generalization.
4. **Series**: convergence tests; power series; radius of convergence.

Example (ε-δ proof). Prove $\lim_{x\to 2} x^2 = 4$. Let $\varepsilon > 0$ be given. We need $\delta > 0$ such that $0 < |x - 2| < \delta \implies |x^2 - 4| < \varepsilon$. Factor: $|x^2 - 4| = |x - 2|\,|x + 2|$. If we first require $\delta \le 1$, then $|x - 2| < 1$ implies $1 < x < 3$, so $|x + 2| < 5$. Choose $\delta = \min\{1, \varepsilon/5\}$. Then $|x - 2| < \delta$ gives $|x^2 - 4| < 5 \cdot (\varepsilon/5) = \varepsilon$. This is the standard pattern: bound $|x - a|$ to control $|f(x) - L|$.

For review, be able to: write ε-N and ε-δ proofs; prove convergence/divergence of sequences and series; determine uniform versus pointwise convergence; apply IVT and EVT; define and work with Riemann integrals; prove properties of continuous functions on compact sets. Identify the ε bound, the uniform modulus of continuity, the dominating function for convergence theorems, and the partition refinement for integration.

8.2 Mathematical spine

$$\lim_{n\to\infty} a_n = L \iff \forall \varepsilon > 0\ \exists N\ \forall n \ge N:\ |a_n - L| < \varepsilon$$

$$\lim_{x\to a} f(x) = L \iff \forall \varepsilon > 0\ \exists \delta > 0\ \forall x:\ 0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon$$

$$\int_a^b f(x)\,dx = \lim_{\|\Delta\|\to 0} \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i$$

Section summary Real analysis places calculus on rigorous foundations using ε-δ arguments. Key concepts: convergence of sequences and series (Cauchy criterion, Bolzano-Weierstrass), continuity (IVT, EVT, uniform continuity), Riemann integration (FTC, integrability conditions), and function space convergence (pointwise vs. uniform). This material develops the logical infrastructure underpinning all of calculus.

9 Algebra: Groups, Rings, Fields, and Symmetry

9.1 Core ideas

Abstract algebra studies algebraic structures. A **group** $(G, *)$ is a set with a binary operation satisfying: closure, associativity $(a * b) * c = a * (b * c)$, identity $e * a = a * e = a$, and inverses $a * a^{-1} = a^{-1} * a = e$. A group is **abelian** if it is commutative ($a * b = b * a$). Examples: $(\mathbb{Z}, +)$, $(\mathbb{R} \setminus \{0\}, \times)$, the symmetric group S_n (permutations of n elements), $GL_n(\mathbb{R})$ (invertible $n \times n$ matrices). A **subgroup** $H \leq G$ is a subset closed under the operation and inverses. The **order** of a group is its cardinality; the order of element a is the smallest n with $a^n = e$.

A **homomorphism** $\phi : G \to H$ preserves the operation: $\phi(a * b) = \phi(a) * \phi(b)$. The **kernel** $\ker \phi = \{g \mid \phi(g) = e_H\}$ is a normal subgroup; the image $\operatorname{im} \phi$ is a subgroup of H. An **isomorphism** is a bijective homomorphism; isomorphic groups are structurally identical. **Cosets** of subgroup H are $gH = \{gh \mid h \in H\}$; they partition G. Lagrange's theorem: $|H|$ divides $|G|$. A **normal subgroup** $N \trianglelefteq G$ satisfies $gNg^{-1} = N$ for all g, allowing the **quotient group** G/N to be formed. The **First Isomorphism Theorem**: $G/\ker \phi \cong \operatorname{im} \phi$.

A **ring** $(R, +, \cdot)$ has two operations: $(R, +)$ is an abelian group, multiplication is associative and distributes over addition. Examples: $(\mathbb{Z}, +, \cdot)$, polynomial ring $\mathbb{R}[x]$, matrix ring $M_n(\mathbb{R})$. A **field** is a commutative ring where nonzero elements have multiplicative inverses. Examples: $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}_p$ for prime p. Polynomials over a field form a Euclidean domain with division algorithm and unique factorization. The **Fundamental Theorem of Algebra**: every nonconstant polynomial in $\mathbb{C}[x]$ has a root in $\mathbb{C}$.

Group actions formalize symmetry. A **group action** $G \times X \to X$ satisfies $e \cdot x = x$ and $(gh) \cdot x = g \cdot (h \cdot x)$. The orbit $Gx = \{gx \mid g \in G\}$ and stabilizer $G_x = \{g \mid gx = x\}$; the orbit-stabilizer theorem: $|G| = |Gx| \cdot |G_x|$. Applications include counting (Burnside's lemma) and classifying symmetries of geometric objects.

1. **Group**: closure, associativity, identity, inverses; subgroup, cosets, Lagrange.
2. **Homomorphism**: kernel, image, isomorphism theorems; quotient groups.
3. **Rings and fields**: polynomial rings, Euclidean domain, $\mathbb{Z}_p$.
4. **Group actions**: orbit, stabilizer, Burnside's lemma.

For review, be able to: verify group axioms; find subgroups and cosets; compute in S_n; construct quotient groups; apply isomorphism theorems; factor polynomials over fields; use group actions to count configurations. Identify the group operation and identity, the kernel of a homomorphism, the field characteristic, and the orbit decomposition.

9.2 Mathematical spine

$$\phi(ab) = \phi(a)\phi(b), \qquad |G| = |Gx| \cdot |G_x|, \qquad G/\ker \phi \cong \operatorname{im} \phi$$

Section summary Algebra abstracts arithmetic to structures: groups (symmetry via invertible operations), rings (arithmetic-like structures with addition and multiplication), and fields (where division works). Key results include Lagrange's theorem (divisibility of group orders), isomorphism theorems (structural relationships), and unique factorization in polynomial rings. Group actions connect algebra to geometry and combinatorics.

10 Topology: Connectedness, Compactness, and Invariants of Shape

10.1 Core ideas

Topology studies properties preserved under continuous deformations (stretching, bending, but not tearing or gluing). A **topological space** $(X, \mathcal{T})$ consists of a set X and a collection $\mathcal{T}$ of open subsets satisfying: (i) $\varnothing, X \in \mathcal{T}$; (ii) arbitrary unions of open sets are open; (iii) finite intersections of open sets are open. The complement of an open set is **closed**. The **metric topology** is induced by a metric d: open balls $B_r(x) = \{y \mid d(x, y) < r\}$ form a basis. $\mathbb{R}^n$ with the Euclidean metric is the motivating example.

A function $f : X \to Y$ between topological spaces is **continuous** if preimages of open sets are open (the topological definition). Equivalently for metric spaces, $x_n \to x \implies f(x_n) \to f(x)$. A **homeomorphism** is a bijective continuous function with continuous inverse; homeomorphic spaces are topologically equivalent (they have the same "shape"). For example, a coffee mug and a donut (torus) are homeomorphic.

Connectedness: X is **connected** if it cannot be written as a disjoint union of two nonempty open subsets. Equivalently, the only clopen (both open and closed) subsets are $\varnothing$ and X. A subset of $\mathbb{R}$ is connected iff it is an interval. The continuous image of a connected set is connected (generalizing IVT). **Path-connectedness** (every pair of points joined by a continuous path) implies connectedness; the converse is false in general. A space is **totally disconnected** if its only connected subsets are singletons (e.g., $\mathbb{Q}$).

Compactness: X is **compact** if every open cover has a finite subcover. In $\mathbb{R}^n$, the Heine-Borel theorem says K is compact iff it is closed and bounded. The continuous image of a compact set is compact. A continuous function on a compact set attains its maximum and minimum (Extreme Value Theorem). Compactness also implies uniform continuity on metric spaces. Sequential compactness (every sequence has a convergent subsequence) is equivalent to compactness in metric spaces.

Homotopy and the **fundamental group** $\pi_1(X, x_0)$ classify spaces up to homotopy equivalence. A **homotopy** between maps $f, g : X \to Y$ is a continuous $H : X \times [0, 1] \to Y$ with $H(x, 0) = f(x)$, $H(x, 1) = g(x)$. Two spaces are **homotopy equivalent** if there exist maps $f : X \to Y$, $g : Y \to X$ with $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$. The fundamental group consists of homotopy classes of loops based at x_0. For the circle S^1, $\pi_1(S^1) \cong \mathbb{Z}$; for S^n $(n \geq 2)$, $\pi_1(S^n)$ is trivial. Simply connected spaces have trivial fundamental group. The **Euler characteristic** $\chi = V - E + F$ for polyhedra is a topological invariant; for a sphere $\chi = 2$, for a torus $\chi = 0$.

1. **Topological space**: open sets, continuity, homeomorphism.
2. **Connectedness**: connected $\iff$ no nontrivial clopen sets; path-connectedness.
3. **Compactness**: Heine-Borel in $\mathbb{R}^n$; extreme value theorem; uniform continuity.
4. **Algebraic topology**: fundamental group π_1, homotopy, Euler characteristic.

For review, be able to: determine if a set is open in a given topology; prove continuity via inverse images of open sets; check connectedness and path-connectedness; apply Heine-Borel for compactness arguments; compute fundamental groups of simple spaces; understand homotopy equivalence. Identify the open cover, the connected component, the compact subset, and the loop homotopy class.

10.2 Mathematical spine

$$f \text{ continuous} \iff f^{-1}(U) \text{ open for all open } U$$

$$\text{Compact: } \forall \text{ open cover } \{U_\alpha\},\ \exists \text{ finite subcover}$$

$$\pi_1(S^1) \cong \mathbb{Z}, \qquad \chi = V - E + F$$

Section summary Topology abstracts notions of proximity and shape. A topological space defines continuity via open sets. Connectedness captures "being all in one piece"; compactness generalizes finiteness and guarantees extreme values and uniform continuity. Algebraic topology uses the fundamental group and Euler characteristic to classify spaces up to deformation, providing invariants that distinguish spheres, tori, and other manifolds.

11 Numerical Mathematics: Stability, Matrix Decompositions, and Optimization

11.1 Core ideas

Numerical mathematics develops algorithms for continuous mathematical problems with guaranteed accuracy. **Conditioning** measures how much a problem's output changes relative to input perturbations. The **condition number** $\kappa = \sup_{\delta x}(\|\delta f\|/\|f\|)/(\|\delta x\|/\|x\|)$; problems with large κ are ill-conditioned. **Numerical stability** means the algorithm produces nearly the exact output for nearly correct input (i.e., backward stable). **Forward error** $= \|\tilde{x} - x\|$; **backward error** = smallest Δx such that $f(x + \Delta x) = \tilde{x}$.

Floating-point arithmetic approximates real numbers with finite precision machine epsilon $\epsilon_{\text{mach}} \approx 2^{-52} \approx 2.22 \times 10^{-16}$ in double precision. Rounding produces relative error $\approx \epsilon_{\text{mach}}$. Catastrophic cancellation occurs when subtracting nearly equal numbers. Algorithms should avoid such cancellation and use numerically stable formulas.

Matrix decompositions are the workhorses of numerical linear algebra. **LU decomposition** $PA = LU$ (permutation, lower triangular, upper triangular) solves linear systems $Ax = b$ in $O(n^3)$ time via forward/backward substitution. **Cholesky decomposition** $A = LL^T$ applies to symmetric positive definite matrices (twice as fast). **QR decomposition** $A = QR$ (orthogonal Q, upper triangular R) provides stable least-squares solutions. **Singular value decomposition** $A = U\Sigma V^T$ gives the optimal low-rank approximation (Eckart-Young theorem): truncating to rank k minimizes $\|A - A_k\|_F$ over rank-k matrices. **Eigenvalue computation** uses the QR algorithm (iterative QR on shifted A).

Iterative methods solve large sparse systems: **Jacobi iteration** $x^{(k+1)} = D^{-1}(b - (L + U)x^{(k)})$; **Gauss-Seidel** uses latest components; **conjugate gradient** for SPD matrices minimizes error in the energy norm. The **condition number** of A affects convergence rate.

Numerical optimization finds minima of functions. **Gradient descent** $x_{k+1} = x_k - \alpha\nabla f(x_k)$ converges linearly for well-conditioned problems. **Newton's method** $x_{k+1} = x_k - H_f^{-1}(x_k)\nabla f(x_k)$ converges quadratically near the solution but requires Hessian inversion. **Quasi-Newton methods** (BFGS) approximate the Hessian. For constrained optimization, **Lagrange multipliers** and **KKT conditions** characterize optimality. **Linear programming** minimizes $c^T x$ subject to $Ax \leq b$, $x \geq 0$ via the simplex method or interior-point methods.

Numerical integration (quadrature) approximates $\int_a^b f(x)\,dx$: trapezoidal rule has error $O(h^2)$, Simpson's rule $O(h^4)$, Gaussian quadrature $O(h^{2n})$. **Adaptive quadrature** refines subintervals where the function varies rapidly. **Differential equation solvers**: Euler's method (first-order), Runge-Kutta methods (classical RK4 is fourth-order). Stiff equations require implicit methods (backward Euler, BDF) for stability.

1. **Stability**: condition number, forward/backward error, machine epsilon.
2. **Decompositions**: LU, Cholesky, QR, SVD; QR algorithm for eigenvalues.
3. **Iterative methods**: Jacobi, Gauss-Seidel, conjugate gradient.

4. **Optimization**: gradient descent, Newton, BFGS, KKT, linear programming.
5. **Quadrature & ODEs**: trapezoidal, Simpson, RK4, adaptive methods.

For review, be able to: compute condition number and interpret it; perform and apply LU, QR, and SVD factorizations; derive convergence criteria for iterative solvers; set up gradient descent and Newton iterations; understand error bounds for quadrature rules; identify stiffness in ODEs. Identify the source of ill-conditioning, the best matrix factorization for the problem, the convergence rate of iterative methods, and the dominant error term in approximation.

11.2 Mathematical spine

$$A = LU, \qquad A = QR, \qquad A = U\Sigma V^T$$
$$x_{k+1} = x_k - \alpha \nabla f(x_k), \qquad x_{k+1} = x_k - H_f^{-1}(x_k)\nabla f(x_k)$$

Section summary Numerical mathematics bridges continuous mathematics and computational implementation. Key themes: conditioning and stability ensure reliable computation; matrix decompositions (LU, QR, SVD) provide the algorithmic backbone for linear algebra; iterative methods and optimization algorithms solve large-scale problems; quadrature and ODE solvers approximate continuous processes. The interplay of mathematical insight and algorithmic design produces efficient, accurate, and stable numerical methods.