PhD Bootcamp: Calculus and Analysis

Rick Presman

Slides adapted from Zeki Kazan, Andrea Aveni, and Bo Liu

Department of Statistical Science

August 14, 2023

Limit

- **Limit of a Sequence:**

$$\lim_{n\to\infty} a_n = a,$$

if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that for all $n > N$, $|a_n - a| < \epsilon$.

- **Limit of a Function:**

$$\lim_{x\to x_0} f(x) = a,$$

if for any $\epsilon > 0$, there exists $\delta > 0$, such that for all $|x - x_0| < \delta$, $|f(x) - a| < \epsilon$.

Finding limits I

1. Guess and show:
$$\lim_{n\to\infty} \frac{1}{n} \sin n.$$
2. Plug in (requires continuity):
$$\lim_{x\to 3} x^2 + 2x.$$
3. Compare order:
$$\lim_{n\to\infty} \frac{5n^2 + 3n + 1}{3n^2 + 6n}.$$

Finding limits II

1. Squeeze theorem:
$$\lim_{x\to 0} \frac{\sin x}{x}.$$
2. Monotone bounded:
$$a_{n+1} = \frac{1+a_n}{2}, a_0 = 5. \lim_{n\to\infty} a_n.$$
3. Other methods.

Important limit

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

Existence of limits (sequence)

- Limits of sequences may not exist, but limit inferior and limit superior must exist.

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty}\left(\inf_{m\geq n} a_n\right), \quad \limsup_{n\to\infty} a_n = \lim_{n\to\infty}\left(\sup_{m\geq n} a_n\right).$$

- Limits exist only when limit inferior and limit superior coincide.
- Show that the limit of $\cos n$ does not exist.

Existence of limits (function)

- The limit $\lim_{x \to x_0} f(x)$ exists if and only if $\lim_{n \to \infty} f(a_n)$ exists (and they must equal) for all sequence $a_n \to x_0$.
- At which points does the limit exist for the Riemann function defined on $[0, 1]$?

$$R(x) = \begin{cases} 1/q, & \text{if } x = p/q \text{ with } p \in \mathbb{Z},\ q \in \mathbb{N} \text{ co-prime}, \\ 0, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Series

- For sequence $\{a_n\}$, we can define an induced sequence $\{S_n\}$ such that

$$S_n = \sum_{k=1}^{n} a_k.$$

- If $\lim_{n\to\infty} S_n$ exists, we say that the series

$$\sum_{k=1}^{\infty} a_k$$

converges to that limit.

- Otherwise, we say that the series diverges.

Geometric Series

- **Finite Geometric Series:** If $r \neq 1$, then

$$\sum_{n=0}^{N-1} ar^n = \frac{a(1-r^N)}{1-r}$$

- **Infinite Geometric Series:** If $-1 < r < 1$, then

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

p-Series

- **p-Series:** The series $\sum_{n=1}^{\infty} \frac{1}{n^p}$
  - Diverges if $p \leq 1$
  - Converges if $p > 1$
- **Harmonic Series:** In the important special case where $p = 1$,

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

  - *Fun Fact:* It can be shown that $\sum_{n=1}^{k} \frac{1}{n} = \ln(k) + \gamma + \mathcal{O}(1/k)$.

Convergence of Series

- If $\sum |a_n|$ converges, then $\sum a_n$ converges.
- If $|a_n| \geq |b_n|$, then

$$\sum_{n=1}^{\infty} a_n \text{ converges} \Longrightarrow \sum_{n=1}^{\infty} b_n \text{ converges.}$$

- If $a_n > 0$ is monotonically decreasing, then

$$\sum_{n=1}^{\infty} (-1)^n a_n$$

converges.

Continuity

- $f(x)$ **is continuous at** x_0: for all $\epsilon > 0$, there exists $\delta > 0$, such that

$$|f(x) - f(x_0)| < \epsilon, \quad \forall |x - x_0| < \delta.$$

- $f(x)$ **is uniformly continuous**: the choice of δ does not depend on x_0. (δ is uniform across all x_0 values.)

Differentiation

- $f(x)$ **is differentiable at** x_0:

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

  exists.
- The limit $f'(x_0)$ is said to be the **derivative** of f evaluated at x_0.
- Differentiability implies continuity.

Function series

- We can define the finite sum of function sequence $\{f_n(x)\}$:

$$S_n(x) = \sum_{k=1}^{n} f_k(x).$$

- We define the series

$$S(x) = \lim_{n\to\infty} S_n(x) = \sum_{k=1}^{\infty} f_k(x),$$

if the sequence converges pointwise for each x.

Function series (cont'd)

- The convergence of a function sequence is actually based upon the existence of the limit

$$\lim_{n\to\infty} \sum_{k=1}^{n} f_k(x),$$

which can be written by the $\epsilon - N$ language.

- If the N is uniform across all x, we say that the convergence is uniform.

Convergence of Function Series

Weierstrass M-test: If

- $|f_n(x)| \leq M_n$,
- $\sum M_n$ converges,

then the series

$$\sum_{n=1}^{\infty} f_n(x)$$

converges absolutely and uniformly.

Power Series

- A power series
$$\sum_{n=0}^{\infty} a_n(x-c)^n$$
converges for all $|x-c| < r$ and diverges for all $|x-c| > r$.
- r is called the **radius of convergence** for the power series.
- **(Cauchy-Hadamard)**
$$r^{-1} = \limsup_{n\to\infty} \sqrt[n]{|a_n|}.$$
- The convergence is absolute whenever $|x-c| < r$ and uniform on any compact set within $(c-r, c+r)$.

Taylor Series

- For an infinitely differentiable function f, the Taylor series about 0 is

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = f(0) + f'(0)x + \frac{f''(0)}{2} x^2 + \mathcal{O}(x^3).$$

- The Taylor series for e^x is, for all $x \in \mathbb{R}$,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \mathcal{O}(x^4).$$

Integral (Sketch on whiteboard)

- Intuitively, the integral of a function is the (signed) area between the function and x-axis.
- If we cut the area into many small slices, we can bound the area of each slice.
- This process gives the lower-bound and upper-bound of the area.
- The integral exists if both bounds can be equal.

p-Integrals

- If $p > -1$, then

$$\int_0^1 x^p \, \mathrm{d}x = \frac{1}{p+1}$$

Otherwise, the integral diverges.

- If $p < -1$, then

$$\int_1^\infty x^p \, \mathrm{d}x = -\frac{1}{p+1}$$

Otherwise, the integral diverges.

Gamma Function

- **Gamma Function:** The Gamma function, Γ, is defined to be

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\,\mathrm{d}x$$

- For $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$
- For all $z > 0$, $\Gamma(z+1) = z\Gamma(z)$
- For $z = \frac{1}{2}$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma'(1) = -\gamma$, where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

Gamma Function II

- The below integral arises often in statistics, for $a, b \in \mathbb{R}^+$

$$\int_0^\infty x^{a-1} e^{-bx}\,\mathrm{d}x = \frac{\Gamma(a)}{b^a}$$

- **Gamma Distribution:** $X \sim \text{Gamma}(\alpha, \beta)$, for $\alpha, \beta \in \mathbb{R}^+$, if X has PDF

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Beta Function

- **Beta Function:** The Beta function, B, is defined to be

$$\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$$

- Relationship to Gamma function: $\mathrm{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$
- **Beta Distribution:** $X \sim \text{Beta}(\alpha, \beta)$, for $\alpha, \beta \in \mathbb{R}^+$, if X has PDF

$$f(x) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Kernels

- **Kernel of a Distribution:** Form of the PDF or PMF of a distribution with factors that are not functions of any variables of the domain omitted.
- Often, tricky-looking integrals and sums encountered in statistics can be easily computed by recognizing the integral as the kernel of a known distribution.
- *Example:* Finding the expectation of $X \sim \text{Exp}(\lambda)$.

$$E[X] = \int_0^\infty x\,\lambda e^{-\lambda x} dx = \lambda \int_0^\infty \underbrace{x^{2-1} e^{-\lambda x}}_{\text{Gamma}(2,\lambda)} dx = \lambda \cdot \frac{\Gamma(2)}{\lambda^2} = \frac{1}{\lambda}$$

Differentiation under the Integral Sign

Under certain regularity assumptions on f, we have that

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

This can be useful in dealing with difficult integrals. For instance,

$$\begin{aligned}\int_0^\infty e^{-t}\ln(t)\,dt &= \int_0^\infty \frac{\partial}{\partial x}\left[e^{-t}t^{x-1}\big|_{x=1}\right]\,dt \\ &= \frac{\partial}{\partial x}\left[\int_0^\infty e^{-t}t^{x-1}\,dt\right]\bigg|_{x=1} \\ &= \Gamma'(1) = -\gamma\end{aligned}$$

(Recall that $t^{x-1} = e^{(x-1)\ln(t)}$)

Differentiation under the Integral Sign

The same theorem applies with summations. For example, we have that (for $|x| < 1$):

$$\begin{aligned}
\sum_{n=0}^{\infty} nx^n &= x \sum_{n=0}^{\infty} nx^{n-1} \\
&= x \sum_{n=0}^{\infty} \frac{\partial}{\partial x}[x^n] \\
&= x \frac{\partial}{\partial x}\left[\frac{1}{1-x}\right] \\
&= \frac{x}{(1-x)^2}
\end{aligned}$$

Gradients and Jacobians

- Gradients extend (first) derivatives in a natural way to multivariate functions $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}^m$:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right)^\intercal$$

- If $f = (f_1, \ldots, f_n)$ with $f_i : \Omega \to \mathbb{R}^n$, then the gradient of f is often referred to as the Jacobian:

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{pmatrix}$$

- Note that sometimes Jacobian refers to the *determinant* of the Jacobian *matrix*

Some Useful Facts

Often we have to deal with taking gradients of expressions with matrices. This can be tedious to do component-wise. Luckily, there are a few tricks:

- $f(\mathbf{x}) = \mathbf{a}^\intercal \mathbf{x} \implies \nabla_{\mathbf{x}} f(\mathbf{x}) = \mathbf{a}$
- $f(\mathbf{x}) = \mathbf{x}^\intercal \mathbf{A} \mathbf{x} \implies \nabla_{\mathbf{x}} f(\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\intercal)\mathbf{x}$

An incredibly useful reference is **The Matrix Cookbook**.

Infinite Sets

Here we consider the concept of cardinality, that is how big a set is. We denote the cardinality of a set A by $|A|$.

Definition

Two sets A, B have the same cardinality ($|A| = |B|$) if there exists a bijective function from A to B.

- This is an equivalence relation on the class of all sets (verify it!) and the cardinal numbers are precisely the equivalence classes of this relation.
- If A is finite, then $|A| = n$ for some $n \in \mathbb{N}$. If A, is infinite, then there are many possibilities since not all the infinities are the same.

Famous Cardinals

- The smallest infinite cardinal number is denoted $\aleph_0$ and is the cardinality of $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. Sets of this cardinality are called **countable**. Any larger set is said to be **uncountable**.
- The second largest cardinal (assuming AC) is denoted $\aleph_1$ and (assuming CH) is equal to $2^{\aleph_0}$, that is, it is the cardinality of the power set of any countable set. This is the cardinality of $\mathbb{R}$, $\mathbb{R}^{27}$, the set of all Borel-measurable sets and the set of all integer sequences. This cardinality is sometimes called **the cardinality of the continuum** $\mathfrak{c}$.
- The next cardinal is (assuming the GCH) $\aleph_2 = 2^{\aleph_1}$ and this is the cardinality of the collection of all Lebesgue-measurable sets and of all real functions.

Why is that $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$?

To prove that $|\mathbb{N}| = |\mathbb{Z}|$, we need to find a bijection ϕ between them. An example is

n	0	1	2	3	4	5	6	...	n	...
$\phi(n)$	0	1	-1	2	-2	3	-3	...	$(-1)^{n+1}\lfloor (n+1)/2 \rfloor$	...

Now, to show that $|\mathbb{N}| = |\mathbb{Q}|$, we need to show that there exists some $f : \mathbb{Q} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{Q}$ both surjective. For f we can take a function that associate to a rational number its denominator when reduced in lowest terms. As for g we can take the following spiral function:

A surjective function $g : \mathbb{N} \to \mathbb{Z}^2$

$$
\begin{tikzcd}
(2,-2) \arrow[d] & (2,-1) \arrow[l] & (2,0) \arrow[l] & (2,1) \arrow[l] & (2,2) \arrow[l] \\
\cdots & (1,-1) \arrow[d] & (1,0) \arrow[l] & (1,1) \arrow[l] & (1,2) \arrow[u] \\
 & (0,-1) \arrow[d] & (0,0) \arrow[r] & (1,0) \arrow[u] & (0,2) \arrow[u] \\
 & (-1,-1) \arrow[r] & (-1,0) \arrow[r] & (-1,1) \arrow[r] & (-1,2) \arrow[u]
\end{tikzcd}
$$

Why is that $|\mathbb{N}| \neq |\mathbb{R}|$?

This is the first proof that there are different infinities and was found by Cantor (who got mad studying the infinity) in 1891.

Assume we can list all the real number in (just) $(0, 1]$ written in base two like this

n	x_n
1	0.10101001010111...
2	0.11000100010000...
3	0.10100010101011...
4	0.00101100101001...
$\vdots$	$\vdots$

Then, if we consider the number obtained by "inverting" the n^{th} digit of the n^{th} number (in this example 0.0001...), we would get to conclude that this number is not in the list.

Metric Spaces

Definition

A set X, together with a function $d : X \times X \to \mathbb{R}$ is a **metric space** *if*

1. $d(x, y) = 0$ *iff* $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, z) \leq d(x, y) + d(y, z)$ *(triangle inequality)*

- *Fun fact:* Dropping condition 1 yields a pseudo-metric space
- *Fun fact:* Dropping condition 2 yields a semi-metric space

Convergence

Definition

A sequence x_n in a metric space is said to converge to a point x, if

$$\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N} : \forall n \geq N_\varepsilon, \ d(x_n, x) \leq \varepsilon$$

That is, for any arbitrary small radius, ε, all but finitely many points of the sequence are at distance less than ε from x.

Definition

A sequence x_n is Cauchy if

$$\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N} : \forall n, m \geq N_\varepsilon, \ d(x_n, x_m) \leq \varepsilon$$

Completeness

Definition

A metric space where every Cauchy sequence converges is a **complete** *metric space.*

The idea is that complete metric spaces are metric spaces without holes.

- The space $\mathbb{Q}$ with the usual metric, is not complete. Indeed, we can consider

$$x_n = \frac{(1+\sqrt{2})^{n+1} - (1-\sqrt{2})^{n+1}}{(1+\sqrt{2})^n - (1-\sqrt{2})^n}$$

This sequence is made of rational numbers, it is Cauchy but it does not converge in $\mathbb{Q}$. (Its limit would be $1+\sqrt{2}$)

Compactness I

Compactness is another property of a metric space that is linked (less trivially) with convergence.

Definition

A space X is **compact** *if for every open cover of X we can find a finite subcover.*

- The open set $(0, 1]$ with the usual metric is not compact. Indeed we can consider the open cover (for let's say $\varepsilon = 1/17$)

$$\mathcal{U} = \left\{ \left(\frac{1-\varepsilon}{2^{n+1}}, \frac{1+\varepsilon}{2^n} \right) : n \in \mathbb{N} \right\}$$

Compactness II

Theorem

Heine-Borel Theorem: *If A is a subset of $\mathbb{R}^n$, it is compact iff it is closed and bounded.*

A compact set must be closed and bounded, but the reverse might be incorrect.
Consider $X = \mathbb{N}$ with discrete metric

$$d(x, y) = \mathbb{I}_{\{x \neq y\}}.$$

Convexity I

Definition

A subset A of a vector space is **convex** *if the segment connecting any two points in A lies completely in A.*

$$\forall x, y \in A, \forall \alpha \in [0,1], \;\; \alpha x + (1-\alpha)y \in A$$

Definition

A function $f : A \subseteq V \to \mathbb{R}$ from a convex set to the reals, is **convex** *if*

$$\forall x, y \in A, \forall \alpha \in [0,1] \;\; f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Convexity II

Definition

A function $f : A \subseteq V \to \mathbb{R}$ from a convex set to the reals, is **strictly convex** *if*

$$\forall x, y \in A, \forall \alpha \in (0,1), \;\; f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y)$$

- A twice-differentiable function, f, of a single variable is convex iff $\forall x,\ f''(x) \geq 0$ (strictly if $\forall x\ f''(x) > 0$).
- If f is convex on an open convex subset of $\mathbb{R}^n$, then f is continuous.

What is Measure Theory?

- It is a rigorous (i.e. mathematical) way to associate real numbers to subsets of $\mathbb{R}$, $\mathbb{R}^n$ or any interesting reference space X.
- The numbers measure “how big” the subset is.
- Probability theory is a special case of measure theory.

Algebras

Definition

If X is a set, then an **algebra** *on X, is a collection of subsets $\mathcal{A} \subseteq 2^X$ s.t.*

- $X \in \mathcal{A}$
- $\mathcal{A}$ *is closed under complements, i.e.*

$$\forall A \in \mathcal{A}, A^c \in \mathcal{A}$$

- $\mathcal{A}$ *is closed under finite unions, i.e.*

$$\forall A, B \in \mathcal{A}, A \cup B \in \mathcal{A}$$

σ-Algebras

Definition

An algebra $\mathcal{A}$ is a σ-algebra, if it is closed under **countable** *unions,*

$$(\forall n \in \mathbb{N}, A_n \in \mathcal{A}) \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

Examples

- $\mathcal{A} = \{\emptyset, X\}$ *is the smallest σ-algebra on any X.*
- $\mathcal{A} = 2^X$ *is the largest σ-algebra on any X.*
- $\mathcal{A} = \{A \subset \mathbb{R} : |A| < \infty \vee |A^c| < \infty\}$ *is only an algebra.*

How to generate (σ-)algebras

- Let $\mathcal{C}$ be any collection of subsets. Is there any (σ-)algebra containing $\mathcal{C}$?
  - YEP! The (σ-)algebra 2^X always works!
- If $\mathcal{A}_\alpha$ is a (σ-)algebra $\forall \alpha \in I$, is $\bigcap_{\alpha \in I} \mathcal{A}_\alpha$ a (σ-)algebra? YEP!
  - $\forall \alpha \in I$, $X \in \mathcal{A}_\alpha$, then $X \in \bigcap_{\alpha \in I} \mathcal{A}_\alpha$
  - If $A \in \mathcal{A}_\alpha$ $\forall \alpha \in I$, then $\forall \alpha \in I$ $X^C \in \mathcal{A}_\alpha$ and so $A^C \in \bigcap_{\alpha \in I} \mathcal{A}_\alpha$
  - The same works for finite or countable unions.

Therefore we can define the operator that gives the smallest (σ-)algebra containing any collection $\mathcal{C}$. We use the notation

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{B} \text{ is } \sigma\text{-algebra containing } \mathcal{C}} \mathcal{B}, \quad \mathcal{A}(\mathcal{C}) = \bigcap_{\mathcal{B} \text{ is algebra containing } \mathcal{C}} \mathcal{B}$$

The Nice σ-Algebras

- If X is a metric (or topological) space, then there is a natural σ-Algebra, the **Borel σ-algebra** generated by the open sets.

$$\mathcal{B} = \sigma(\mathcal{T})$$

  - This is the most natural σ-algebra on $\mathbb{R}^n$ and its subsets.
  - It contains quite a lot of sets ($|\mathcal{B}_{\mathbb{R}}| = 2^{\aleph_0}$) but not all of them ($|2^{\mathbb{R}}| = 2^{2^{\aleph_0}}$).
  - It is impossible to "explicitly" construct a non-Borel set, so don't worry, you (probably) won't ever encounter a non-Borel set in your life.

Measures

Definition

A **measure** *is is a function* $\mu : \mathcal{A} \to \overline{\mathbb{R}}_+$ *from an algebra to the non-negative numbers that satisfies*

- *Non-negativity:* $\forall A \in \mathcal{A}$, $\mu(A) \geq 0$
- *Null Empty Set:* $\mu(\emptyset) = 0$
- σ*-additivity: If* $(A_n)_{n \in \mathbb{N}}$ *are disjoint measurable sets such that their union is in* $\mathcal{A}$, *then*

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

Special Kinds of Measures

Definition

A measure μ on $\mathcal{A}$ is σ-finite if there are sets $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{A}$ s.t.

$$X = \bigcup_{n\in\mathbb{N}} A_n \;\wedge\; \forall n \in \mathbb{N},\, \mu(A_n) < \infty$$

Definition

A measure μ on $\mathcal{A}$ is finite if $\mu(X) < \infty$

Definition

A measure μ on $\mathcal{A}$ is a **probability** *if $\mu(X) = 1$*

Examples

- We can consider $\mathbb{R}$ with the full σ-algebra $\mathcal{A} = 2^{\mathbb{R}}$. Then a nice measure on this space is the **counting measure** $\mu(A) = |A|$.
  - On an uncountable set (like $\mathbb{R}$), this is not σ-finite.
- If X is any set with any σ-algebra and $x \in X$, then the **Dirac measure** δ_x is defined as

$$\delta_x(A) = \begin{cases} 1 & \textit{if } x \in A \\ 0 & \textit{if } x \notin A \end{cases}$$

  - The Dirac measure is always a probability measure.
- Let $X = \{1, 2, 3\}$ and let $\mathcal{A} = 2^X$, then a unique measure μ can be specified by imposing $\mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = 1/3$.