PhD Bootcamp: Calculus and Analysis

Zeki Kazan and Andrea Aveni

Department of Statistical Science

August 11, 2021

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

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

Taylor Series

- For an arbitrary 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 $x \in \mathbb{R}$,

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

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 $\sum_{n=1}^{k} \frac{1}{n} = \ln(k) + \gamma + \mathcal{O}(1/k)$.

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 $X \sim \text{Exp}(\lambda)$.

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

Limits

- **Definition of e^x:** e^x can be expressed via the following limit

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

- **L'Hospital's Rule:** If $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ or $\pm\infty$, then

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}$$

  - *Technical Sidenote:* f and g must be differentiable on an open interval, I, except possibly at $c \in I$

Completing the Square (One Dimension)

- **Completing the Square:** A technique for converting an equation of the form $ax^2 + bx + c$ to the form $a(x-h)^2 + k$
- *Example:* For $a = 3$, $b = 12$, and $c = 27$,

$$\begin{aligned} 3x^2 + 12x + 27 &= 3(x^2 + 4x + 9) \\ &= 3(x^2 + 4x + 4 + 5) \\ &= 3(x+2)^2 + 15 \end{aligned}$$

- In general, $h = -\frac{b}{2a}$ and $k = c - \frac{b^2}{4a}$

Differentiation under the Integral Sign

Under certain regularity assumptions of f, we have that

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

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

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

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

Differentiation under the Summation Sign

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

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

Calculus Exercises I

1. Let $X \sim \text{Beta}(\alpha, \beta)$. Find $E[X]$ and simplify as much as possible.
2. Evaluate the following limit in TWO different ways (using methods covered in these slides)

$$\lim_{x \to 0} \frac{e^x - (1 + x)}{x^2}$$

3. Let $X \sim \text{Exp}(\lambda)$. Find a general form for $E[X^k]$.
4. Identify which of the following series and integrals converge

$$1.\ \int_0^1 x^2\,\mathrm{d}x \qquad 3.\ \int_1^\infty x^2\,\mathrm{d}x \qquad 5.\ \sum_{n=1}^\infty n^2$$

$$2.\ \int_0^1 x^{-2}\,\mathrm{d}x \qquad 4.\ \int_1^\infty x^{-2}\,\mathrm{d}x \qquad 6.\ \sum_{n=1}^\infty n^{-2}$$

Calculus Exercises II

5. X has a two-sided geometric distribution with parameter p if it has PMF $\mathrm{P}[X = x] = cp^{|x|}$ for $x \in \mathbb{Z}$
   1. Find the normalizing constant, c
   2. Find the CDF, $F(x) = \mathrm{P}[X \leq x]$
   3. Find $E[X]$ and $Var[X]$ (*Hint:* Find $\sum_{n=0}^{\infty} n^2 x^n$)
6. Let $X \sim \mathcal{N}(0, 1)$ and $Y \mid X \sim \mathcal{N}(X, 1)$. Determine the distribution of Y.
   - *Hint:* Recall that $f_Y(y) = \int f_{Y|X}(y|x) f_X(x)\, \mathrm{d}x$
   - Recall that $Z \sim \mathcal{N}(\mu, \sigma^2)$ has PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(z-\mu)^2\right\}$$

Feel free to reach out to me at zekican.kazan@duke.edu with questions or to check your solutions.

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 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 $g : \mathbb{Z} \to \mathbb{Q}$ 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 $g : \mathbb{Z} \to \mathbb{Z}^2$

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

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

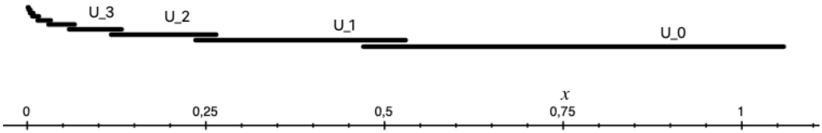

Figure 1: Illustration of our chosen open covering of $(0, 1]$.

You can see and you should try to prove that this is indeed an open covering (meaning that for any $x \in (0, 1]$, there exists some $U \in \mathcal{U}$ such that $x \in U$) and that it does not admit any finite subcover. (Hint: consider the numbers of the form $3/2^{n+2}$ and prove that they belong to only one $U \in \mathcal{U}$.)

Compactness III

Theorem

(Assuming X is a metric space) If $A \subseteq X$ is compact, then it is closed.
If A is a closed subset of a compact set X, then, A is compact.

Theorem

A compact set A in a metric space X is bounded.

Theorem

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

Compactness and Convergence

Theorem

If A is a subset of a metric space, than A is compact iff every sequence in A admits a convergent subsequence.

The real numbers are complete (they are the completion of $\mathbb{Q}$), but $\mathbb{R}$ is not compact.
For instance, the sequence $x_n = n$ does not admit any convergent subsequence, or analogously, the cover $\mathcal{U} = \{(n - 1/31, n + 1/31) : n \in \mathbb{Z}\}$ does not admit any finite subcover.

Continuity I

Definition

A function $f : X \to Y$ on metric spaces is continuous *if*

$$\forall x \in X, \forall \varepsilon > 0, \exists \delta_{x,\varepsilon} > 0 : f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$$

Definition

A function $f : X \to Y$ on metric spaces is uniformly continuous *if*

$$\forall \varepsilon > 0, \exists \delta_\varepsilon > 0, \forall x \in X, : f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$$

- *Recall:* An open ball of radius $\varepsilon > 0$ centered at a point x is $B_\varepsilon(x) := \{y \in X : d(x, y) < \varepsilon\}$

Continuity II

- If $x_n \to x$ and f is continuous, then $f(x_n) \to f(x)$.
- The continuous image of a compact set is compact.
- Heine-Cantor: A continuous function on a compact set is uniformly continuous.
- Extreme Value Theorem: If $f : X \mapsto \mathbb{R}$ is continuous and X is compact, then f attains its maximum and minimum.
- If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then it is continuous.

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*

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

- 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.

Convexity III

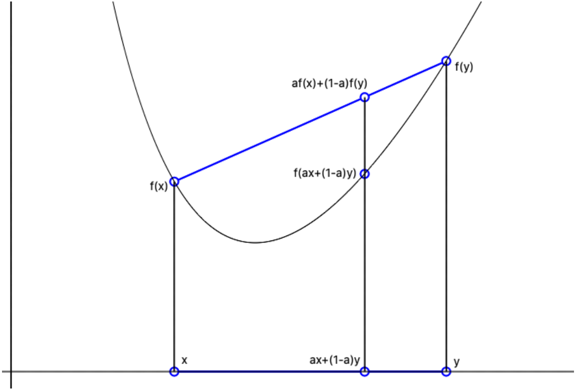

Figure 2: A convex function

What is Measure Theory?

- It is a rigorous (i.e. mathematical) way to associate real numbers to subsets of $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^\omega$ 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.*
- $\mathcal{A} = \{A \subset \mathbb{R} : |A| \leq \aleph_0 \vee |A^c| \leq \aleph_0\}$ *is a σ-algebra.*
- $\mathcal{A} = \{A \subset \mathbb{R} : |A| < \infty \vee |A^c| < \infty\}$ *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 $\bigcap_{\alpha \in I} A_\alpha$ a (σ-)algebra? YEP!
  - If $X \in A_\alpha \ \forall \alpha \in I$, then $X \in \bigcap_{\alpha \in I} A_\alpha$
  - If $X \in A_\alpha \ \forall \alpha \in I$, then $X^C \in A_\alpha \ \forall \alpha \in I$ and so $X^C \in \bigcap_{\alpha \in I} A_\alpha$
  - 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*

1. *Non-negativity:* $\forall A \in \mathcal{A},\ \mu(A) \geq 0$
2. *Null Empty Set:* $\mu(\emptyset) = 0$
3. σ*-additivity: If* $(A_n)_{n\in\mathbb{N}}$ *are disjoint 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

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

  - 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 $\mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = 1/3$.

Analysis Exercises

1. Verify that the open coverings in slides 24 and 27 do not admit any finite subcover.
2. Find a σ-algebra $\mathcal{A}$ with 8 elements.
3. Prove that $\sigma(\{x\} : x \in \mathbb{R}) \subsetneq \mathcal{B}_{\mathbb{R}}$.
4. Let $\mathcal{A}$ be the finite-cofinite σ-algebra on $\mathbb{Z}$, and let μ be such that $\mu(A) = 1$ if A is infinite and 0 otherwise, prove that this is not a measure.
5. Find a set in $\mathcal{B}_{\mathbb{R}}$ that is not in $\sigma(\{(-x, x) : x > 0\})$
6. What are the functions $f : \mathbb{R} \to \mathbb{R}$ such that both f and $-f$ are convex?
7. *Challenge:* Show that $\mathcal{B}_{\mathbb{R}} = \sigma(\{[a, b) \cap \mathbb{R} : a, b \in \overline{\mathbb{Q}} \wedge b > a\})$.

Send your solutions/questions to andrea.aveni@duke.edu