Metric Spaces

Motivation

In a metric space, attention is focused on the spatial relationships between elements. In the previous lecture, we began building the framework for social science theory by thinking about how elements in a set are ordinally ranked. This is clearly a necessary mathematical concept for building theories occupied by actors with preferences. Distance is a second notion that we will find useful in theory building. For example, we often talk about polarization in political science. We need a concept of distance to analyze how far apart the ideal policy of two legislators or parties are. An important large class of models in political science are called "spatial models" that build on this intuition and formalize utility as a decreasing function of the distance between an agent's ideal outcome and the realized outcome of a political process. This lecture introduces the mathematical concepts necessary to formalize and apply the concept of distance in political science research.

Definition

A **metric space** is a set X on which is defined a measure of distance between the elements. To conform with our conventional notion of distance, the distance measure must satisfy certain properties. The distance should be positive. It should be symmetric: the distance from Princeton to Cambridge should equal the distance between Cambridge and Princeton.

Last, the shortest route between two distinct elements should be the direct route. A distance measure with these properties is called a "metric."

Formally, a **metric** on a set X is a measure that associates with every pair of points $x, y \in X$ a real number $d(x, y)$ satisfying the following properties:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x)$ (symmetry)
4. $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality)

A metric space is a pair (X, d) where X is a set and d is a metric. Elements of a metric space are usually called **points**.

Example: The most familiar metric space is $(\mathbb{R}, d)$ where $d(x, y) = |x - y|$.

Example: Consider how we might define the distance between policy bundles represented as points in $\mathbb{R}^n$. Given a two dimensional policy space (social and economic policy for example), one way to measure the distance between two policy pairs x and y is to consider the difference in each dimension of policy and sum them. That is

$$d_1(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

An alternative would be to square the differences and take their square root:

$$d_2(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

We may also consider that the policy whose position has changed the most should determine the distance between policy pairs:

$$d_\infty(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

The metric d_2 is common and referred to as the **Euclidean metric** which generalizes the

usual notion of distance in two and three dimensional space. Remember the Pythagorean theorem? The metric d_∞ is known as the "supremum metric." ■

In practice we will often use objects called "metric subspaces." If X is a metric space and Y is a nonempty subset of X, we can view Y as a metric space in its own right using the distance function induced by d on Y, $d|_{Y\times Y}$. We say that $(Y, d|_{Y\times Y})$ is a **metric subspace** of X. For example, $[0,1]$ can be through of as a metric subspace of $\mathbb{R}$ where the distance between $x, y \in [0,1]$ is calculated by viewing x and y as points in $\mathbb{R} : d_1(x,y) = |x-y|$. We may of course also think of $[0,1]$ as a metric subspace of $\mathbb{R}^2$. Formally, we identify $[0,1]$ with $[0,1] \times \{u\}$ for any $u \in \mathbb{R}$. This renders the distance between x and y equal again to $|x-y|$ e.g. $d_2((x,u),(y,u)) = |x-y|$.

Open and Closed Sets

We want to have a concept of proximity when we use metric spaces. The set of points in close proximity to a given point x_0 is called a "ball" about x_0. Formally, an **open ball** about x_0 with radius r is the set of points

$$B_r(x_0) = \{x \in X : d(x, x_0) < r\}$$

Open balls are not necessarily spherical: their shape depends on the metric.

A set $S \subseteq X$ is called a **neighborhood** of x_0 if S contains an open ball about x_0 ($\exists B_\epsilon(x_0) \subset S$ for some $\epsilon > 0$). A point $x_0 \in S$ is called an **interior point** of S if S contains an epsilon ball about x_0. An open ball is a symmetrical neighborhood but the concept of neighborhood does not require symmetry. The set of interior points of a set S is called the **interior** of S, denoted int S. A set S is **open** if $S = \text{int } S$. A point $x_0 \in X$ is called a **boundary point** of $S \subseteq X$ if every neighborhood of x_0 contains points of S and S^c. The boundary $b(S)$ is the set of all boundary points of S. The **closure** $\overline{S}$ of S is the union of S with its boundary, i.e. $\overline{S} = S \cup b(S)$. A set is **closed** if $S = \overline{S}$.

Example: The boundary of the unit ball $B_1(0)$ is the set $S_1(0) = \{x \in X : d(x, 0) = 1\}$. This is called the "unit sphere." In $\mathbb{R}^2$ the unit sphere is $S_1(0) = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$, which is the boundary of the set $B_1(0) = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}$. ■

Convergence

A **sequence** in a metric space X is a list of particular elements $x^1, x^2, ...$ of X. Let (x^n) denote a sequence. We often encounter sequences in political science. A sequence of observations of some political variable such as presidential approval is a "time series." The moves in a game theoretic model are also a sequence. In your quantitative analysis courses, you will often analyze the asymptotic properties of estimators by appealing to properties of a sequence of estimates of a population parameter.

A sequence (x^n) **converges** to x if for all $r > 0$ there exists an N such that $x^n \in B_r(x)$ for all $n \geq N$. Equivalently, (x^n) converges to x if for all $\epsilon > 0$, there exists some N such that for all $n \geq N$, $d(x^n, x) < \epsilon$. We often denote convergence by $x^n \to x$ or $\lim x^n = x$.

Example: Let $(x^n) = \frac{3n+1}{7n-4}$ be a sequence in $(\mathbb{R}, d_1)$. This sequence converges to 3/7. To see this, let $\epsilon > 0$ and let $N = \frac{19}{49\epsilon} + \frac{4}{7}$. Then $n > N$ implies $n > \frac{19}{49\epsilon} + \frac{4}{7}$, hence $\frac{19}{7(7n-4)} = \frac{3n+1}{7n-4} - \frac{3}{7} < \epsilon$. Therefore for $n \geq N$, $|\frac{3n+1}{7n-4} - \frac{3}{7}| < \epsilon$. ■

If a sequence converges, it has a unique limit. To see this, assume $x^n \to x$ and $x^n \to y$. By the triangle equality, $d(x, y) \leq d(x, x^n) + d(x^n, y)$ for all n. $d(x, x^n) \to 0$ and $d(x^n, y) \to 0$, $d(x, y) \to 0$. This implies $x = y$.

A sequence is bounded if there exists some L such that $d(x^n, y) \leq L$ for all n where $L \in \mathbb{R}$ and y is a point in X. Every convergent sequence is bounded. Proving this is left as an exercise.

Sequences allow an alternative characterization of closed sets from the one given above.

Proposition 1 *A set S in a metric space X is closed if and only if every sequence in S that converges in X converges to a point in S.*

Proof:

First we prove that closed implies that the limit of a convergent sequence is in S by contradiction. Let S be a closed subset of X and take (x^n) a sequence in S with $x^n \to x$ for some $x \in X$. If $x \in X \backslash S$ then we can find an $r > 0$ such that $B_r(x) \subseteq X \backslash S$ because $X \backslash S$ is open in X. But since $d(x^n, x) \to 0$, there must exist an N such that $x^N \in B_r(x)$. This contradicts the assertion that all terms of the sequence (x^m) lie in S.

Now we prove that convergence to a point in S implies that S is closed by contraposition. Suppose S is not closed in X. Then $X \backslash S$ is not open. Therefore we can find an $x \in X \backslash S$ such that every open ball around x intersects S. Thus for any $n = 1, 2, ...$, there is an $x^n \in B_{\frac{1}{n}}(x) \cap S$. But then $(x^n) \in S$ and $\lim x^n = x$ yet $x \notin S$. Thus were S is not closed, there would exist at least one sequence in S that converges to a point outside of S. ■

We will also occasionally use **subsequences** to prove properties of sequences. A subsequence of (x^n) is a sequence (y^k) where $y^k = x^{n_k}$ where $n_1 < n_2 < ...$ is an increasing sequence of indices. We often denote a subsequence of (x^n) with (x^{n_k}).

Example: Let $(x^n) = 1, 1/2, 1/3, 1/4, 1/5, ...$ The sequence $(y^k) = 1, 1/3, 1/5, ...$ is a subsequence of (x^n) where $n_1 = 1$, $n_2 = 3$, $n_3 = 5$, etc. ■

Compactness

Compactness is one of the most fundamental concepts in real analysis and plays an important role in optimization theory. You've probably encountered a definition of compact that is loosely something like "a set that is closed and bounded." This is true in some metric spaces but is not a proper definition of compact. We need an additional definition to properly define compactness.

Let X be a metric space and $S \subseteq X$. A class $\mathcal{O}$ of subsets of X is said to **cover** S if $S \subseteq \cup\mathcal{O}$. If all members of such a class $\mathcal{O}$ are open in X, then we say that $\mathcal{O}$ is an **open cover**. A subset of $\mathcal{O}$ that also covers S is called a **subcover**. Now we are ready to define

compactness.

A metric space X is **compact** if every open cover of X has a finite subset that also covers X.

Example: Consider the interval $(0,1)$ (a subset of $\mathbb{R}$) and the collection $\mathcal{O} = \{(\frac{1}{i}, 1) : i = 1, 2, ...\}$. Note that $(0,1) = (\frac{1}{2}, 1) \cup (\frac{1}{3}, 1) \cup ....$ Therefore $\mathcal{O}$ is an open cover of $(0,1)$. However, $\mathcal{O}$ does not have a finite subset that covers $\mathcal{O}$. Therefore $(0,1)$ is not compact. ■

Example: $[0,1]$ is a compact subset of $\mathbb{R}$. To see this, suppose there exists an open cover $\mathcal{O}$ of $[0,1]$ such that no finite subset of $\mathcal{O}$ covers $[0,1]$. This implies that either $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$ is not covered by any finite subset of $\mathcal{O}$. Pick any one of the intervals with this property and call it $[a_1, b_1]$. Then either $[a_1, \frac{1}{2}(a_1 + b_1)]$ or $[\frac{1}{2}(b_1, +a_1), b_1]$ is not covered by any finite subset of $\mathcal{O}$. Pick any one of these intervals with this property and call it $[a_2, b_2]$. Continue this inductively to obtain two sequences, (a_m) and (b_m) in $[0,1]$ such that

(i) $a_m \leq a_{m+1} < b_{m+1} \leq b_m$,

(ii) $b_m - a_m = \frac{1}{2^m}$,

(iii) $[a_m, b_m]$ is not covered by any finite subset of $\mathcal{O}$,

for each $m = 1, 2, ...$ Properties (i) and (ii) allow us to find a real number c with $\{c\} = \cap^{\infty}[a_i, b_i]$. Now take any $O \in \mathcal{O}$ which contains c. Since O is open and $\lim a_m = \lim b_m = c$, we must have $[a_m, b_m] \subset O$ for a large enough m. But this contradicts condition (iii). We conclude that $[0,1]$ is compact. ■

Of course, we are familiar with a simpler version of compactness. This is primarily due to the fact that we typically work in $\mathbb{R}^n$ with a Euclidean metric. The **Heine-Borel Theorem** is to thank for this result. The theorem establishes that the following two statements are equivalent for a subset S of $(\mathbb{R}^n, d_2)$:

1. S is closed and bounded
2. S is compact, i.e. every open cover of S has a finite subcover

In general, compactness implies closed and bounded but the reverse is not guaranteed to hold.

An additional pair of facts will help us find other compact sets. First, any closed subset of a compact metric space is compact. Second, the product of two compact metric spaces is compact (Tychonoff's theorem). This latter fact will be very useful in game theory when we want to determine whether a strategy space $S = S_1 \times S_2 \times ... \times S_n$ is compact. If each player $i = 1, 2, ...n$ has a compact strategy space, the strategy space is compact. A compact strategy space will be a key component of the fixed point theorems we use to prove the existence of a Nash equilibrium.

Cauchy Sequences

A sequence (x^n) in a metric space X is called a **Cauchy sequence** if for any $\epsilon > 0$, there exists an $N \in \mathbb{R}$ such that $d(x^k, x^l) < \epsilon$ for all $k, l \geq N$.

Example: $(x^n) = 1/2^n$ is a Cauchy sequence in $(\mathbb{R}, d_1)$. Take $\epsilon > 0$ and choose N so large that $2^{-N} < \epsilon$. Then if $n, m > N$, we have

$$|x^n - x^m| = |1/2^n - 1/2^m| \leq 1/2^n + 1/2^m \leq 1/2^N + 1/2^N < \epsilon \quad \blacksquare$$

Cauchy sequences have several properties:

1) If (x^n) is convergent, then it is Cauchy.

2) If (x^n) is Cauchy, then (x^n) is bounded but need not converge in X.

3) If (x^n) is Cauchy and has a subsequence that converges in X, then (x^n) converges in X as well.

Proving Property 1 and the first part of Property 2 are left as exercises. To see the second part of Property 2, consider the sequence $(x^n) = (1, 1/2, 1/3, ...)$ in $(0, 1]$. This sequence is Cauchy but does not converge in this space (while it does in $[0, 1]$ or $\mathbb{R}$). For Property 3, say that (x^n) is Cauchy and has a convergent subsequence $(x^{n_k}) \to x$. This implies

$$d(x^n, x) \leq d(x^n, x^{n_k}) + d(x^{n_k}, x) \to 0$$

as $n, k \to \infty$.

These properties can be quite useful. Say we are given a sequence (x^n) in some metric space and need to check if the sequence converges. Doing this requires guessing a limit x and then showing that the sequence actually converges to x. This is not always efficient. An alternative approach is simply to check whether the sequence is Cauchy. If not Cauchy, we conclude by Property 2 that the sequence is not convergent. If it is Cauchy, however, we still can't say whether or not it converges because of Property 2, illustrated in the counterexample above. If we knew something about the metric space or the metric space it is a subspace of, may learn something. For example, what if we knew that in our metric space, all Cauchy sequences converged? It turns out that most of the metric spaces we will use in fact have this property!

Completeness

Compactness generalizes the concept of finiteness and is one of two fundamental properties of metric spaces. The other property, **completeness**, generalizes the idea of richness. An incomplete space lacks certain necessary elements.

A **complete metric space** is defined as a metric space X in which every Cauchy sequence in X converges to a point in X.

Example: We saw above that $(0, 1]$ with the d_1 metric is not complete. ■

It is natural to think about a relationship between completeness and closedness.

Proposition 2 *Let X be a metric space and Y a metric subspace of X. If Y is complete, then it is closed in X. Conversely, if Y is closed in X and X is complete, then Y is complete.*

Proof: Let Y be complete and take any $(x^n) \in Y$ that converges in X. Because (x^n) converges, it is Cauchy and thus $\lim x^n \in Y$. We saw above that if a sequence in Y that converges in X converges to a point in Y, then Y is closed.

Now assume X is complete and Y is closed in X. If (x^n) is a Cauchy sequence in Y, the by the completeness of X, it must converge in X. But since Y is closed, $\lim x^n$ must be an element of Y. Therefore Y is complete. ■

A corollary of this is that a metric subspace of a complete metric space is complete if and only if it is closed in X.

One final fact about completeness will be useful: every compact metric space is complete.

Intro to Fixed Point Theory

Why do we care about completeness? It turns out that certain types of mappings on a complete metric space possess a very desirable "fixed point" property. A **self-map** is a function, $f(\cdot)$, whose domain and codomain are identical. A **fixed point** of a self-map is a point that satisfies the property $f(x) = x$. Fixed points will be extremely useful in political science theory for ensuring that a model has an equilibrium.

One type of self-map is called a **contraction**. Let X be a metric space. A self-map f on X is a contraction if there exists a real number $K \in (0, 1)$ such that

$$d(f(x), f(y)) \leq Kd(x, y)$$

for all $x, y \in X$.

Example: Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = \frac{x}{2}$. To verify that this is a contraction, consider two arbitrary points, $x, y \in \mathbb{R}$. Also note that in general, $a|b| \leq |ab|$. We need to show that $d(f(x), f(y)) \leq Kd(x, y)$ for some $K \in (0, 1)$. $d(f(x), f(y)) = |f(x) - f(y)| = |\frac{x}{2} - \frac{y}{2}| = |\frac{1}{2}(x - y)| \leq \frac{1}{2}|x - y| = \frac{1}{2}d(x, y)$. Since $\frac{1}{2} \in (0, 1)$, f is indeed a contraction. ■

When a contraction is defined on a complete metric space, a contraction must map a point to itself. In other words, it *must* have a fixed point. In fact, we can show that it must have a *unique* fixed point!

Theorem 1 (Contraction Mapping) *if X is a complete metric space and f is a contrac-*

tion defined on X, then there exists a unique $x^* \in X$ such that $f(x^*) = x^*$.

Proof: Let (X, d) be a complete metric space and let $f : X \to X$ be a contraction. To show the existence of a fixed point for f, pick any $x \in X$ and define $(x^n) \in X^\infty$ such that $x^{n+1} = f(x^n)$ for $n = 0, 1, ...$ This is a Cauchy sequence: let K be the contraction coefficient of f and notice that $d(x^{n+1}, x^n) \leq K^n d(x^1, x^0)$ for all $n = 1, 2, ...$ To see this, note that by construction, $d(x^{n+1}, x^{n+2}) \leq K d(x^n, x^{n+1}) \leq K^2 d(x^{n-1}, x^n) \leq ... \leq K^n d(x^0, x^1)$. Thus for any $k > l + 1$,

$$d(x^k, x^l) \leq d(x^k, x^{k-1}) + ... + d(x^{l+1}, x^l)$$

$$\leq (K^{k-1} + ... + K^l) d(x^1, x^0)$$

$$= \frac{K^l(1 - K^{k-l})}{1 - K} d(x^1, x^0)$$

Therefore $d(x^k, x^l) < \frac{K^l}{1-K} d(x^1, x^0)$.

This implies that (x^n) is a Cauchy sequence. Since X is complete, our Cauchy sequence (x^n) must converge to some $x^* \in X$.

Then for any $\epsilon > 0$, there must exists some N such that $d(x^*, x^n) < \frac{\epsilon}{2}$ for all $n = N, N + 1, ...,$

and therefore

$$d(f(x^*), x^*) \leq d(f(x^*), x^{n+1}) + d(x^{n+1}, x^*)$$

$$= d(f(x^*), f(x^n)) + d(x^{n+1}, x^*)$$

$$\leq K d(x^*, x^n) + d(x^{n+1}, x^*)$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2}$$

Since $\epsilon > 0$ is arbitrary, we must have $d(f(x^*), x^*) = 0$ which is possible only if $f(x^*) = x^*$.

To prove uniqueness, note that if $x \in X$ was another fixed point, we would have $d(x, x^*) = d(f(x), f(x^*)) \leq K d(x, x^*)$ which is possible only if $x = x^*$. ■

The Contraction Mapping Theorem is also commonly referred to as the **Banach Fixed Point Theorem**.

Exercises

1) Show that $d_\infty(x, y) = \max_{i=1}^{n} |x_i - y_i|$ is a metric for $\mathbb{R}^n$.

2) Consider the metric space $(\mathbb{R}, d_1)$ and the sequence $(\frac{1}{n^2})$. Prove that $\lim \frac{1}{n^2} = 0$.

3) Prove that every convergent sequence in a metric space is bounded.

4) Prove that every convergent sequence in a metric space is Cauchy.

5) Prove that every Cauchy sequence is bounded.

6) The Bolzano-Weierstrass theorem states that every bounded sequence of real numbers has a convergent subsequence. Use the theorem to prove that $\mathbb{R}$ is complete.

Source Material

Carter, M. (2001). *Foundations of Mathematical Economics.* Cambridge, MA: The MIT Press.

Ok, E.A. (2007). *Real Analysis with Economic Applications.* Princeton, NJ: Princeton University Press.

Ross, K.A. (2013). *Elementary Analysis: The Theory of Calculus.* New York, NY: Springer.