Metric Entropy

SDS 391P.6, Spring 2026

Pratik Patil

These notes are a work in progress and are provided as-is for instructional purposes only. They are not (yet) at the level of a scholarly document. In particular, the notes draw from various sources and do not (yet) have sufficient references to the original sources. Additionally, almost surely the notes have errors and they are only probably approximately correct. The notes will be updated regularly as the course progresses. Last updated: 2026-02-26.

1 Motivation

A recurring theme in our next module is that high-dimensional random objects are often controlled by *suprema*:

$$Y = \sup_{t \in T} X_t,$$

where $(X_t)_{t \in T}$ is a random process indexed by a set T. For example, the operator norm of a matrix $A \in \mathbb{R}^{m \times n}$ can be written as

$$\|A\| = \sup_{x \in S^{n-1}} \|Ax\|_2, \qquad S^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}.$$

Suprema like this are ubiquitous in high-dimensional statistics and probability. In covariance estimation, for instance, one wants to control

$$\|\widehat{\Sigma} - \Sigma\| = \sup_{x \in S^{d-1}} \left| x^\top (\widehat{\Sigma} - \Sigma) x \right|,$$

which again is a supremum over the sphere.[1]

A basic way to attack a supremum is to *discretize* the index set T. The language for discretization is:

- ϵ-*nets*, which replace an infinite set by finitely many representatives;
- *covering numbers* and *packing numbers*, which quantify how many points we need;
- *metric entropy*, the logarithm of the covering number, which acts as a complexity measure.

A second motivation is geometric: different sets have radically different metric entropy. The Euclidean sphere is "large" at fine scales (exponential in dimension), while many polytopes are "small" (their entropy depends polynomially on the number of vertices). This contrast provides some intuition behind why some high-dimensional problems are feasible: the effective complexity is not always the ambient dimension. Structure greatly helps in high dimensions!

We will develop the definitions and basic bounds in this lecture, and we will study concrete applications in the upcoming lectures.

[1] We will return to this application in the next lecture. For now, we only use it as motivation.

2 Recall: metric spaces

A typical deterministic reduction is:

- Rewrite an object of interest as a supremum over a metric space (T, d).
- Build a finite ϵ-net $\mathcal{N} \subset T$.
- Bound $\sup_{t \in T} X_t$ by $\max_{t \in \mathcal{N}} X_t$ plus a discretization error.

The reason this matters probabilistically is that $\max_{t \in \mathcal{N}} X_t$ can often be controlled by a union bound once $|\mathcal{N}|$ is finite. The price we pay is the *size* of the net, i.e., a covering number.

We therefore start by reviewing the metric notions that organize all such arguments.

A metric space is a pair (T, d) where T is a set and $d : T \times T \to [0, \infty)$ satisfies:

$$d(s,t) = 0 \iff s = t, \qquad d(s,t) = d(t,s), \qquad d(s,u) \leqslant d(s,t) + d(t,u).$$

We write the open ball of radius ϵ around t as

$$B(t, \epsilon) := \{s \in T : \; d(s,t) < \epsilon\}.$$

For $K \subset T$, the diameter is

$$\mathrm{diam}(K) := \sup_{s,t \in K} d(s,t).$$

The main example for us is Euclidean space $(\mathbb{R}^n, \|\cdot\|_2)$.

3 Covering and packing

3.1 Covering numbers

Definition 3.1 (ϵ-net)**.** Let (T, d) be a metric space and $K \subset T$. A set $\mathcal{N} \subset K$ is an ϵ-*net* of K if for all $x \in K$, there exists an $x_0 \in \mathcal{N}$ such that $d(x, x_0) \leqslant \epsilon$. Equivalently, the balls $\{B(x_0, \epsilon) : x_0 \in \mathcal{N}\}$ cover K.

Definition 3.2 (Covering number)**.** The *covering number* $\mathcal{N}(K, d, \epsilon)$ is the smallest cardinality of an ϵ-net of K.

When the metric is clear we write $\mathcal{N}(K, \epsilon)$.

3.2 Packing numbers

Definition 3.3 (Packing number)**.** A set $\mathcal{M} \subset K$ is ϵ-*separated* if $d(x,y) > \epsilon$ for all distinct $x, y \in \mathcal{M}$. The *packing number* $\mathcal{P}(K, d, \epsilon)$ is the maximum cardinality of an ϵ-separated subset of K.

If $\mathcal{M}$ is ϵ-separated, then the closed balls of radius $\epsilon/2$ centered at $\mathcal{M}$ are disjoint (triangle inequality). See Figure 1 for an illustration.

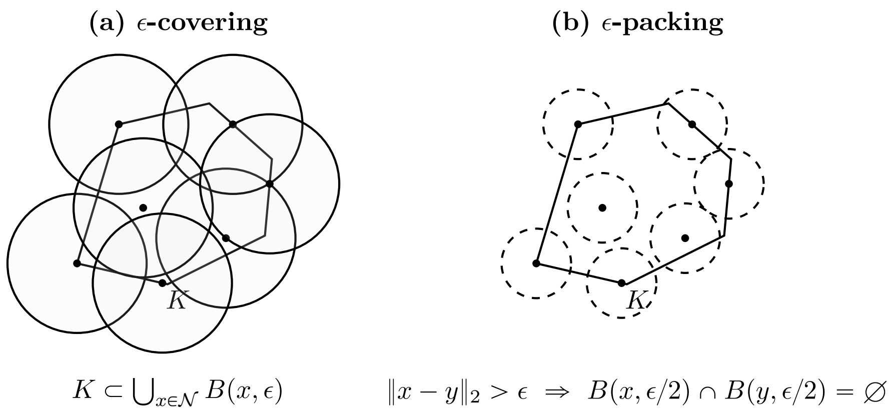

Figure 1: Covering versus packing. Panel (a) shows an ϵ-covering: ϵ-balls centered at points of $\mathcal{N} \subset K$ cover K. Panel (b) shows an ϵ-packing: points in $\mathcal{N}$ are pairwise separated by $> \epsilon$, hence their $(\epsilon/2)$-balls are disjoint. Note that these $(\epsilon/2)$-balls are allowed to extend outside K; only disjointness matters.

3.3 Covering versus packing

Lemma 3.4 (Maximal separated sets give nets)**.** *Let $\mathcal{M} \subset K$ be a maximal ϵ-separated set (i.e. you cannot add any point of K while preserving separation). Then $\mathcal{M}$ is an ϵ-net of K.*

Proof sketch. If some $x \in K$ were farther than ϵ from all points of $\mathcal{M}$, we could add x and keep separation, contradicting maximality. □

Lemma 3.5 (Covering versu packing comparison)**.** *For any $K \subset T$ and $\epsilon > 0$,*

$$\mathcal{P}(K,d,2\epsilon) \;\leqslant\; \mathcal{N}(K,d,\epsilon) \;\leqslant\; \mathcal{P}(K,d,\epsilon).$$

Proof. The upper bound follows from Lemma 3.4: a maximal ϵ-packing is an ϵ-net. For the lower bound, any ϵ-ball can contain at most one point from a 2ϵ-separated set, hence any ϵ-cover must have at least $\mathcal{P}(K,2\epsilon)$ balls. □

Remark 3.6 (Metric entropy)**.** The quantity

$$H(K,d,\epsilon) := \log \mathcal{N}(K,d,\epsilon)$$

is often called the *metric entropy* of K at scale ϵ. The base of the logarithm is usually irrelevant; we use natural log unless stated otherwise.

4 Volumetric bounds for balls

From now on, take $T = \mathbb{R}^n$ with Euclidean metric $d(x,y) = \|x-y\|_2$. Write $B_2^n := \{x : \|x\|_2 \leqslant 1\}$.

For sets $A, B \subset \mathbb{R}^n$, recall the Minkowsi sum is

$$A + B := \{a + b : \; a \in A, \; b \in B\}.$$

Geometrically, $A + \epsilon B_2^n$ is the ϵ-neighborhood of A:

$$A + \epsilon B_2^n = \{x : \; d(x,A) \leqslant \epsilon\}.$$

See Figure 2 for an illustration.

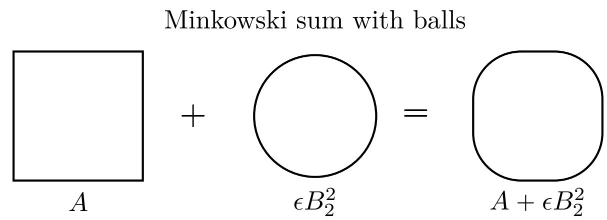

Figure 2: In Euclidean space, adding ϵB_2^n to a set corresponds to taking an ϵ-neighborhood (a "rounded" enlargement).

4.1 A basic volume bound for covering and packing

Proposition 4.1 (Covering and packing versus volumes)**.** *Let $K \subset \mathbb{R}^n$ be measurable and $\epsilon > 0$. Then*

$$\frac{\mathrm{vol}(K)}{\mathrm{vol}(\epsilon B_2^n)} \leqslant \mathcal{N}(K,\epsilon) \leqslant \mathcal{P}(K,\epsilon) \leqslant \frac{\mathrm{vol}\big(K + (\epsilon/2)B_2^n\big)}{\mathrm{vol}((\epsilon/2)B_2^n)}.$$

Proof sketch. If K is covered by N balls of radius ϵ, then $\mathrm{vol}(K) \leqslant N\,\mathrm{vol}(\epsilon B_2^n)$. For the upper bound, take an ϵ-packing $\{x_i\}_{i=1}^M$. The disjoint balls $x_i + (\epsilon/2)B_2^n$ are contained in $K + (\epsilon/2)B_2^n$, so $M\,\mathrm{vol}((\epsilon/2)B_2^n) \leqslant \mathrm{vol}(K + (\epsilon/2)B_2^n)$. □

4.2 Covering numbers of the Euclidean balls

Corollary 4.2 (Euclidean ball)**.** *For $\epsilon > 0$,*

$$\left(\frac{1}{\epsilon}\right)^n \leqslant \mathcal{N}(B_2^n,\epsilon) \leqslant \left(1 + \frac{2}{\epsilon}\right)^n.$$

In particular, for $\epsilon \in (0,1]$,

$$\left(\frac{1}{\epsilon}\right)^n \leqslant \mathcal{N}(B_2^n,\epsilon) \leqslant \left(\frac{3}{\epsilon}\right)^n.$$

The same upper bound holds for the sphere S^{n-1}.

Proof sketch. Use Proposition 4.1 and the scaling $\mathrm{vol}(rB_2^n) = r^n \mathrm{vol}(B_2^n)$. For the sphere, note $S^{n-1} \subset B_2^n$. □

The main takeaway is that metric entropy of the Euclidean ball scales like

$$H(B_2^n,\epsilon) \asymp n\log(1/\epsilon).$$

Even for a *fixed* accuracy ϵ (say $\epsilon = 0.1$), the number of net points is exponential in n.

5 Volumetric bounds for polytopes

Volumetric bounds are often tight for "round" sets like balls and spheres. But for many structured sets, volume bounds are too pessimistic. A particularly important example is a *polytope*, the convex hull of finitely many points.

5.1 Approximate Carathéodory theorem

Definition 5.1 (Convex hull)**.** For $T \subset \mathbb{R}^n$,

$$\operatorname{conv}(T) := \Big\{ \sum_{i=1}^{m} \lambda_i t_i : \ m \in \mathbb{N},\ t_i \in T,\ \lambda_i \geqslant 0,\ \sum_{i=1}^{m} \lambda_i = 1 \Big\}.$$

A polytope is the convex hull of a finite set.

Carathéodory's theorem says that exact representations in $\mathbb{R}^n$ may require $n+1$ points. Surprisingly, *approximate* representations can be dimension-free.

Theorem 5.2 (Approximate Carathéodory)**.** *Assume $T \subset B_2^n$. Then for every $x \in \operatorname{conv}(T)$ and every integer $k \geqslant 1$, there exist $x_1, \dots, x_k \in T$ such that*

$$\left\| x - \frac{1}{k} \sum_{j=1}^{k} x_j \right\|_2 \leqslant \frac{1}{\sqrt{k}}.$$

Proof sketch. Write $x = \sum_{i=1}^{m} \lambda_i t_i$ with $t_i \in T$. Let Z be a random vector taking value t_i with probability λ_i. Then $\mathbb{E}Z = x$ and $\|Z\|_2 \leqslant 1$ almost surely. Let $Z_1, \dots, Z_k$ be i.i.d. copies and set $\bar{Z} = \frac{1}{k} \sum_{j=1}^{k} Z_j$. Independence gives

$$\mathbb{E}\|\bar{Z} - x\|_2^2 = \mathbb{E} \left\| \frac{1}{k} \sum_{j=1}^{k} (Z_j - \mathbb{E}Z_j) \right\|_2^2 = \frac{1}{k^2} \sum_{j=1}^{k} \mathbb{E}\|Z_j - \mathbb{E}Z_j\|_2^2 \leqslant \frac{1}{k}.$$

Thus there exists a realization with $\|\bar{Z} - x\|_2 \leqslant 1/\sqrt{k}$, and $\bar{Z}$ is an average of k points in T. ☐

5.2 Covering numbers for polytopes

Corollary 5.3 (Covering convex hulls of finite sets)**.** *Let $P = \operatorname{conv}\{v_1, \dots, v_N\} \subset B_2^n$. Then for every $k \geqslant 1$,*

$$\mathcal{N}\left(P,\ \frac{1}{\sqrt{k}}\right) \ \leqslant\ N^k.$$

Equivalently, for $\epsilon \in (0, 1]$,

$$H(P, \epsilon) = \log \mathcal{N}(P, \epsilon) \ \leqslant\ C \, \frac{\log N}{\epsilon^2}$$

for absolute constant C.

Proof sketch. Let

$$\mathcal{M}_k := \left\{ \frac{1}{k} \sum_{j=1}^{k} v_{i_j} : \ i_j \in \{1, \dots, N\} \right\}.$$

There are at most N^k such averages. By Theorem 5.2, every $x \in P$ lies within $1/\sqrt{k}$ of some element of $\mathcal{M}_k$, so $\mathcal{M}_k$ is a $(1/\sqrt{k})$-net. ☐

Compare the ball versus polytope:

$$H(B_2^n, \epsilon) \ \asymp\ n \log(1/\epsilon), \qquad H(P, \epsilon) \ \lesssim\ \frac{\log N}{\epsilon^2}.$$

At a fixed accuracy ϵ, the ball needs $\exp(\Theta(n))$ points, while a polytope with N vertices needs at most $N^{O(1)}$ points (dimension-free). This is one precise way to say that many polytopes are "simpler" than spheres.

5.3 Volume bound for polytopes

The entropy bound above has a striking geometric consequence: polytopes with a moderate number of vertices occupy a vanishing fraction of the Euclidean ball in high dimension.

Theorem 5.4 (Carl–Pajor volume bound)**.** *Let $P \subset B_2^n$ be a polytope with N vertices. Then there exists an absolute constant $C > 0$ such that*

$$\frac{\operatorname{vol}(P)}{\operatorname{vol}(B_2^n)} \;\leqslant\; \left(C \sqrt{\frac{\log N}{n}} \right)^n .$$

(When $\log N \gtrsim n$, the right-hand side is $\geqslant 1$, so the inequality is trivial. The interesting regime is $\log N \ll n$.)

Proof sketch. By Corollary 5.3, for each integer $k \geqslant 1$, P can be covered by at most N^k Euclidean balls of radius $1/\sqrt{k}$. Thus

$$\operatorname{vol}(P) \;\leqslant\; N^k \operatorname{vol}\!\left(\frac{1}{\sqrt{k}} B_2^n \right) = N^k\, k^{-n/2} \operatorname{vol}(B_2^n).$$

Choose $k \simeq n/(2\log N)$ (and round to an integer) to optimize $N^k k^{-n/2}$, which yields

$$\frac{\operatorname{vol}(P)}{\operatorname{vol}(B_2^n)} \;\leqslant\; \left(C \sqrt{\frac{\log N}{n}} \right)^n$$

for an absolute constant C. □

Remark 5.5 (A high-dimensional surprise)**.** If $N = \operatorname{poly}(n)$, then $\log N = O(\log n)$ and

$$\left(C \sqrt{\frac{\log N}{n}} \right)^n = \exp\Big(-\Theta(n \log(n/\log n)) \Big),$$

which is astronomically small. In high dimensions, most of the Euclidean ball cannot be captured by a polytope with only polynomially many vertices.

5.4 Balls versus polytopes

It is useful to keep the following scalings in mind:

- Euclidean ball:
$$\mathcal{N}(B_2^n, \epsilon) \;\lesssim\; \left(\frac{C}{\epsilon} \right)^n, \qquad H(B_2^n, \epsilon) \;\lesssim\; n \log(1/\epsilon).$$

- Polytope with N vertices in the unit ball:
$$\mathcal{N}(P, \epsilon) \;\lesssim\; \exp\!\left(C \,\frac{\log N}{\epsilon^2} \right) = N^{C/\epsilon^2}, \qquad H(P, \epsilon) \;\lesssim\; \frac{\log N}{\epsilon^2}.$$

The main takeaway here is that metric entropy can depend on structure (like vertex count), not just ambient dimension.

6 Look ahead

In this lecture we introduced the deterministic geometry behind many probabilistic proofs:

- ϵ-nets discretize complicated index sets.
- Covering and packing numbers quantify the size of nets; their logarithm is metric entropy.
- Volumetric arguments give sharp entropy bounds for Euclidean balls and spheres.
- Maurey's empirical method gives dimension-free entropy bounds for convex hulls of finite sets and shows a sharp contrast between spheres and polytopes (polytopes with few vertices are tiny in high dimension).

Next time, we will use these tools in a concrete statistical setting of covariance estimation. The core move will be to write an operator norm as a supremum over the sphere, discretize by a net, and combine:

$$\text{metric entropy} \;+\; \text{concentration for a fixed direction} \;\Longrightarrow\; \text{uniform control over all directions.}$$

Source material

Parts of this lecture are based on references: Vershynin (2018); Tropp (2023), in addition to the author's accumulated experience working on related topics.

References

Tropp, J. A. (2023). Probability in high dimensions. Caltech CMS Lecture Notes 2021-01.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.