Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 13**

Metric Entropy: Covering and Packing

Lecturer: Chuan-Fa Tang *Scribe: Zijiang Yang*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Motivation

Many statistical problems involve collections of random variables indexed by a set. Examples include a finite collection $X_1, \dots, X_n$, a countably infinite collection $\{X_i : i \in \mathbb{N}\}$, and a stochastic process $\{X_t : t \in [0,1]\}$ with an uncountable index set. Further examples discussed in class are:

- The empirical distribution function $\{F_n(t) : t \in \mathbb{R}\}$.
- Quadratic forms $\{v^\top A v : v \in S^{d-1}\}$, where A is a random matrix and $S^{d-1} = \{v \in \mathbb{R}^d : \|v\|_2 = 1\}$.
- An empirical process $\{P_n f - Pf : f \in \mathcal{F}\}$, where

$$P_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i), \qquad Pf = \mathbb{E}[f(X)].$$

For a finite index set, cardinality provides a natural measure of its size. For an infinite index set, cardinality alone does not describe its geometric complexity. Covering and packing measure this complexity at a given resolution.

2 Metric Spaces and Examples

A metric space is a pair $(\mathbb{T}, \rho)$, where $\mathbb{T}$ is a nonempty set and $\rho : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ satisfies:

1. $\rho(\theta, \theta') \geq 0$, with equality if and only if $\theta = \theta'$.
2. $\rho(\theta, \theta') = \rho(\theta', \theta)$.
3. $\rho(\theta_1, \theta_3) \leq \rho(\theta_1, \theta_2) + \rho(\theta_2, \theta_3)$.

Examples include the following distances, with equation labels matching the textbook [1].

Euclidean distance. On $\mathbb{R}^d$,

$$\rho(\theta, \theta') = \|\theta - \theta'\|_2 = \left(\sum_{j=1}^{d} (\theta_j - \theta'_j)^2\right)^{1/2}. \tag{5.1a}$$

Rescaled Hamming distance. On $\{0,1\}^d$,

$$\rho_H(\theta, \theta') = \frac{1}{d}\sum_{j=1}^{d} \mathbf{1}\{\theta_j \neq \theta'_j\}. \tag{5.1b}$$

This is the fraction of coordinates at which the binary vectors differ.

L^2 distance. On $L^2(\mu, [0,1])$,

$$\rho_2(f,g) = \|f-g\|_2 = \left(\int_0^1 |f(x)-g(x)|^2 \, d\mu(x) \right)^{1/2}. \tag{5.1c}$$

Functions agreeing μ-almost everywhere are identified.

Uniform distance. On $C([0,1])$, the continuous functions on $[0,1]$,

$$\rho_\infty(f,g) = \|f-g\|_\infty = \sup_{x\in[0,1]} |f(x)-g(x)|. \tag{5.1d}$$

3 Covering Numbers and Metric Entropy

For $x \in \mathbb{T}$ and $\delta > 0$, define the closed metric ball

$$B_\rho(x,\delta) = \{y \in \mathbb{T} : \rho(x,y) \le \delta\}.$$

Definition 5.1 (Covering number)**.** A δ-cover of $\mathbb{T}$ with respect to ρ is a set $\{\theta^1, \dots, \theta^N\} \subseteq \mathbb{T}$ such that for every $\theta \in \mathbb{T}$, there is an $i \in \{1, \dots, N\}$ satisfying $\rho(\theta, \theta^i) \le \delta$. Equivalently,

$$\mathbb{T} \subseteq \bigcup_{i=1}^{N} B_\rho(\theta^i, \delta).$$

The δ-covering number $N(\delta; \mathbb{T}, \rho)$ is the cardinality of the smallest δ-cover. If no finite cover exists, it is infinite.

Geometrically, covering asks how few balls of radius δ are needed to cover the set. Packing asks how many sufficiently separated points can be placed in the set; its formal definition will be considered later.

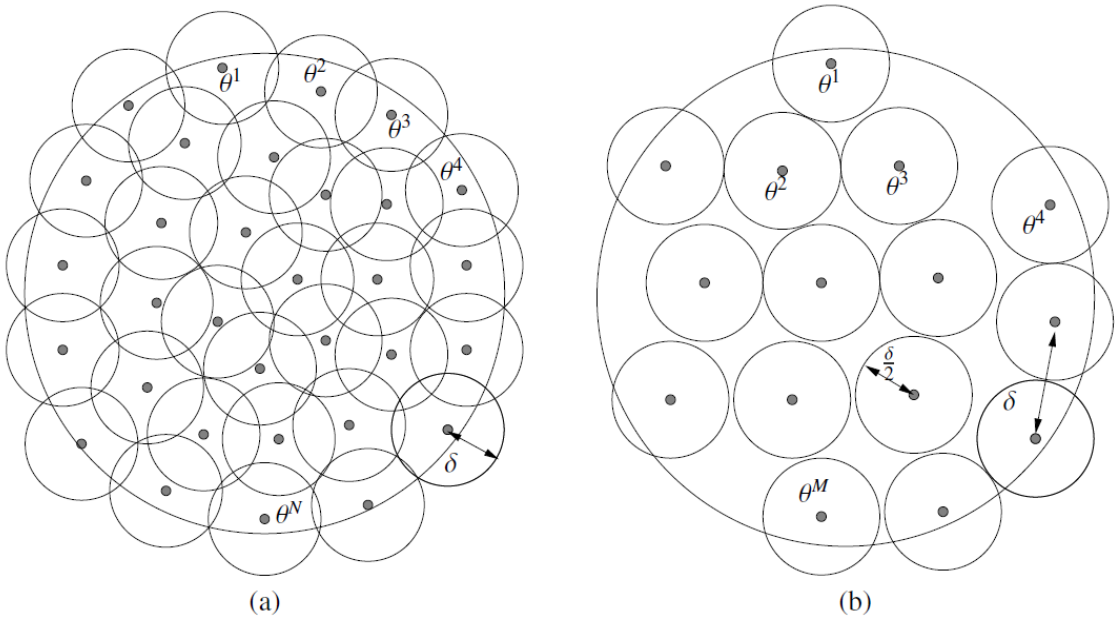


Figure 1: Covering and packing (textbook Figure 5.1). Covering balls have radius δ; the disjoint balls illustrating packing have radius $\delta/2$.

Total Boundedness and Metric Entropy

A metric space is **totally bounded** if $N(\delta; \mathbb{T}, \rho) < \infty$ for every $\delta > 0$. Thus, at every positive resolution, finitely many points suffice to approximate the set.

Example (not totally bounded). Consider the continuous ramp functions drawn in class, with $\mathbb{T} = \{f_j : j \geq 1\}$ and $\rho(f, g) = \|f - g\|_\infty$. Although every function takes values in $[0, 1]$, distinct functions satisfy $\|f_i - f_j\|_\infty > 1/2$. Hence $N(\delta; \mathbb{T}, \rho) = \infty$ for $\delta < 1/4$, so this space is not totally bounded.

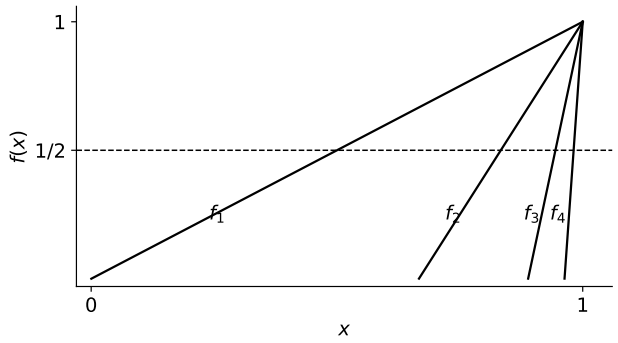


Figure 2: The non-totally bounded example drawn in class.

Covering numbers are non-increasing in the radius:

$$0 < \delta_1 \leq \delta_2 \quad \Longrightarrow \quad N(\delta_1; \mathbb{T}, \rho) \geq N(\delta_2; \mathbb{T}, \rho).$$

The quantity $\log N(\delta; \mathbb{T}, \rho)$ is called the **metric entropy**. It measures the size of the set at resolution δ on a logarithmic scale. Unless a base is specified, log denotes the natural logarithm.

4 Examples of Covering Numbers

Example 5.2 (Covering numbers of unit cubes)**.** Consider $\mathbb{T} = [-1, 1]$ with $\rho(x, y) = |x - y|$. For $\delta > 0$, divide the interval into $\lceil 1/\delta \rceil$ equal subintervals, each of length at most 2δ. Their midpoints form a δ-cover. Consequently,

$$N(\delta; [-1, 1], |\cdot|) \leq \left\lceil \frac{1}{\delta} \right\rceil \leq 1 + \frac{1}{\delta}. \tag{5.2}$$

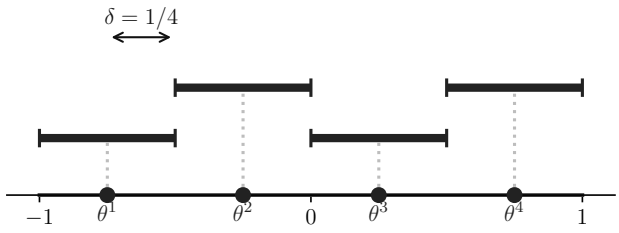
For $[-1, 1]^d$ under the sup-norm, take the Cartesian product of the one-dimensional cover. Each coordinate is approximated within δ, yielding

$$N(\delta; [-1, 1]^d, \|\cdot\|_\infty) \leq \left(1 + \frac{1}{\delta}\right)^d,$$

and hence

$$\log N(\delta; [-1, 1]^d, \|\cdot\|_\infty) \leq d \log \left(1 + \frac{1}{\delta}\right).$$

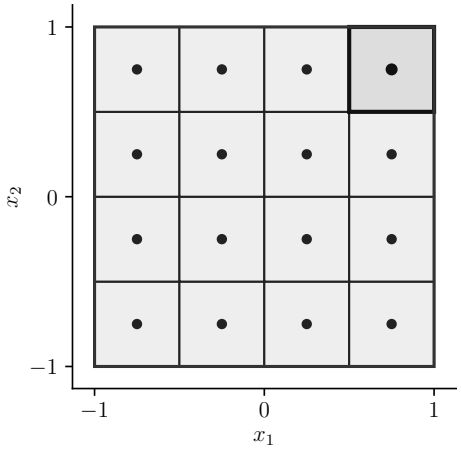
(a) Interval: four balls of radius 1/4



$$[-1,1] \subseteq \bigcup_{i=1}^{4} [\theta^i - \delta, \theta^i + \delta]$$

Dots are centers; segments are the covering balls.

(b) Square: ℓ_∞ balls are squares



$4 \times 4 = 16$ centers, $\delta = 1/4$

Figure 3: Example 5.2: covering $[-1,1]$ and $[-1,1]^2$ with $\delta = 1/4$. In two dimensions, balls under the sup-norm are squares.

Remark 5.3 (Comparison of metrics)**.** For every $x \in \mathbb{R}^d$,

$$\|x\|_\infty \le \|x\|_2 \le \sqrt{d}\,\|x\|_\infty.$$

Therefore, for the same set $\mathbb{T} \subseteq \mathbb{R}^d$,

$$N(\delta; \mathbb{T}, \|\cdot\|_\infty) \le N(\delta; \mathbb{T}, \|\cdot\|_2) \le N(\delta/\sqrt{d}; \mathbb{T}, \|\cdot\|_\infty).$$

This explains the comparison between square and circular covering regions drawn in class. The comparison involves a change in radius; the two covering numbers need not be equal.

Example 5.4 (Covering of the binary hypercube)**.** Consider the binary hypercube $\mathbb{H}^d := \{0,1\}^d$, equipped with the rescaled Hamming metric (5.1b). First, we upper bound its δ-covering number.

For $\delta \in (0,1)$, let

$$S = \{1, 2, \ldots, \lceil (1-\delta)d \rceil\},$$

where $\lceil (1-\delta)d \rceil$ is the smallest integer greater than or equal to $(1-\delta)d$. Consider the set of binary vectors

$$\mathbb{T}(\delta) := \{\theta \in \mathbb{H}^d : \theta_j = 0 \text{ for all } j \notin S\}.$$

For any $\widetilde{\theta} \in \mathbb{H}^d$, construct $\theta \in \mathbb{T}(\delta)$ by matching $\widetilde{\theta}$ on every coordinate in S and setting all coordinates outside S to zero. The vectors agree on S and can differ on at most

$$d - \lceil (1-\delta)d \rceil = \lfloor \delta d \rfloor$$

remaining coordinates. Thus,

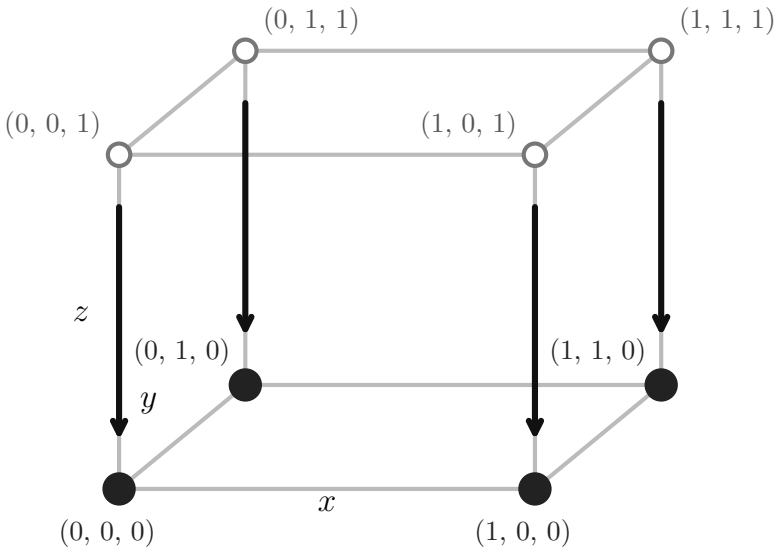
$$\rho_H(\theta, \widetilde{\theta}) \le \frac{\lfloor \delta d \rfloor}{d} \le \delta.$$

This proves that $\mathbb{T}(\delta)$ is a δ-cover. Its coordinates in S can vary freely, so it contains $2^{\lceil (1-\delta)d \rceil}$ vectors. Consequently,

$$N_H(\delta; \mathbb{H}^d) \le 2^{\lceil (1-\delta)d \rceil}, \qquad \text{and hence} \qquad \frac{\log N_H(\delta; \mathbb{H}^d)}{\log 2} \le \lceil d(1-\delta) \rceil.$$

Here $N_H(\delta; \mathbb{H}^d)$ denotes $N(\delta; \mathbb{H}^d, \rho_H)$, as in the textbook.

$$d = 3, \quad \delta = 1/3, \quad m = \lceil (1-\delta)d \rceil = 2$$



$$\mathbb{T}(\delta) = \{(0,0,0), (1,0,0), (0,1,0), (1,1,0)\}$$

Filled dots: cover centers. Arrows: set the third coordinate to zero.

Each mapped vertex differs in one coordinate: $\rho_H = 1/3 = \delta$. Thus $N(\delta) \leq 4$.

Figure 4: Example 5.3 with $d = 3$ and $\delta = 1/3$: retain the first two coordinates and set the third to zero. Four centers cover all eight vertices.

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. Section 5.1, pp. 121–123 (through the upper bound in Example 5.3).