

Introduction to Empirical Processes and Semiparametric Inference Lecture 13: Entropy Calculations

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Uniform Entropy

Let $\mathbf{1}\{\mathcal{C}\}$ denote the collection of all indicator functions of sets in the class \mathcal{C} .

The following theorem gives a bound on the L_r covering numbers of $\mathbf{1}\{\mathcal{C}\}$:

THEOREM 1. *There exists a universal constant $K < \infty$ such that for any VC-class of sets \mathcal{C} , any $r \geq 1$, and any $0 < \epsilon < 1$,*

$$N(\epsilon, \mathbf{1}\{\mathcal{C}\}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$

This is Theorem 2.6.4 of VW, and we omit the proof.

Since $F = 1$ serves as an envelope for $\mathbf{1}\{\mathcal{C}\}$, we have as an immediate corollary that, for

$$\mathcal{F} = \mathbf{1}\{\mathcal{C}\}, \sup_Q N(\epsilon \|F\|_{1,Q}, \mathcal{F}, L_1(Q)) < \infty$$

and

$$J(1, \mathcal{F}, L_2) \lesssim \int_0^1 \sqrt{\log(1/\epsilon)} d\epsilon = \int_0^\infty u^{1/2} e^{-u} du \leq 1,$$

where the supremum is over all finite probability measures Q with $\|F\|_{Q,2} > 0$.

Thus the uniform entropy conditions required in the G-C and Donsker theorems of the previous chapter are satisfied for indicators of VC-classes of sets.

Since the constant 1 serves as a universally applicable envelope function, these classes of indicator functions are therefore G-C and Donsker, provided the requisite measurability conditions hold.

For a function $f : \mathcal{X} \mapsto \mathbb{R}$, the subset of $\mathcal{X} \times \mathbb{R}$ given by $\{(x, t) : t < f(x)\}$ is the *subgraph* of f .

A collection \mathcal{F} of measurable real functions on the sample space \mathcal{X}

- is a *VC-subgraph class* or *VC-class* (for short),
- if the collection of all subgraphs of functions in \mathcal{F} forms a VC-class of sets (as sets in $\mathcal{X} \times \mathbb{R}$).

Let $V(\mathcal{F})$ denote the VC-index of the set of subgraphs of \mathcal{F} .

VC-classes of functions grow at a polynomial rate just like VC-classes of sets:

THEOREM 2. *There exists a universal constant $K < \infty$ such that, for any VC-class of measurable functions \mathcal{F} with integrable envelope F , any $r \geq 1$, any probability measure Q with $\|F\|_{Q,r} > 0$, and any $0 < \epsilon < 1$,*

$$N(\epsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq KV(\mathcal{F})(4e)^{V(\mathcal{F})} \left(\frac{2}{\epsilon}\right)^{r(V(\mathcal{F})-1)}.$$

Thus VC-classes of functions easily satisfy the uniform entropy requirements of the G-C and Donsker theorems of the previous chapter.

A related kind of function class is the *VC-hull* class.

A class of measurable functions \mathcal{G} is a VC-hull class if there exists a VC-class \mathcal{F} such that

- each $f \in \mathcal{G}$ is the pointwise limit of a sequence of functions $\{f_m\}$
- in the *symmetric convex hull* of \mathcal{F} (denoted $\text{sconv}\mathcal{F}$).

A function f is in $\text{sconv}\mathcal{F}$ if $f = \sum_{i=1}^m \alpha_i f_i$, where the α_i s are real numbers satisfying $\sum_{i=1}^m |\alpha_i| \leq 1$ and the f_i s are in \mathcal{F} .

The *convex hull* of a class of functions \mathcal{F} , denoted $\text{conv}\mathcal{F}$, is similarly defined but with the requirement that the α_i 's are positive.

We use $\overline{\text{conv}\mathcal{F}}$ to denote pointwise closure of $\text{conv}\mathcal{F}$ and $\overline{\text{sconv}\mathcal{F}}$ to denote the pointwise closure of $\text{sconv}\mathcal{F}$.

Thus the class of functions \mathcal{F} is a VC-hull class if $\mathcal{F} = \overline{\text{sconv}\mathcal{G}}$ for some VC-class \mathcal{G} .

THEOREM 3. Let Q be a probability measure on $(\mathcal{X}, \mathcal{A})$, and let \mathcal{F} be a class of measurable functions with measurable envelope F , such that $QF^2 < \infty$ and, for $0 < \epsilon < 1$,

$$N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq C \left(\frac{1}{\epsilon}\right)^V,$$

for constants $C, V < \infty$ (possibly dependent on Q).

Then there exist a constant K depending only on V and C such that

$$\log N(\epsilon \|F\|_{Q,2}, \overline{\text{conv}}\mathcal{F}, L_2(Q)) \leq K \left(\frac{1}{\epsilon}\right)^{2V/(V+2)}.$$

This is Theorem 2.6.9 of VW, and we omit the proof.

It is not hard to verify that $\text{sconv}\mathcal{F}$ is a subset of the convex hull of $\mathcal{F} \cup \{-\mathcal{F}\} \cup \{0\}$, where $-\mathcal{F} \equiv \{-f : f \in \mathcal{F}\}$ (see Exercise 9.6.1 below).

Since the covering numbers of $\mathcal{F} \cup \{-\mathcal{F}\} \cup \{0\}$ are at most one plus twice the covering numbers of \mathcal{F} , the conclusion of Theorem 3 also holds if $\overline{\text{conv}}\mathcal{F}$ is replaced with $\overline{\text{sconv}}\mathcal{F}$.

This leads to the following easy corollary for VC-hull classes:

COROLLARY 1. *For any VC-hull class \mathcal{F} of measurable functions and all $0 < \epsilon < 1$,*

$$\sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq K \left(\frac{1}{\epsilon}\right)^{2-2/V}, \quad 0 < \epsilon < 1,$$

where

- *the supremum is taken over all probability measures Q with $\|F\|_{Q,2} > 0$,*
- *V is the VC-index of the VC-subgraph class associated with \mathcal{F} ,*
- *and the constant $K < \infty$ depends only on V .*

We now present several important examples and results about VC-classes of sets and both VC-subgraph and VC-hull classes of functions.

LEMMA 1. *Let \mathcal{F} be a finite-dimensional vector space of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$.*

Then \mathcal{F} is VC-subgraph with $V(\mathcal{F}) \leq \dim(\mathcal{F}) + 2$.

The next three lemmas consist of useful tools for building VC-classes from other VC-classes.

- LEMMA 2. • Let \mathcal{C} and \mathcal{D} be VC-classes of sets in a set \mathcal{X} ,
- with respective VC-indices $V_{\mathcal{C}}$ and $V_{\mathcal{D}}$;
 - and let \mathcal{E} be a VC-class of sets in \mathcal{W} , with VC-index $V_{\mathcal{E}}$.
 - Also let $\phi : \mathcal{X} \mapsto \mathcal{Y}$ and $\psi : \mathcal{Z} \mapsto \mathcal{X}$ be fixed functions.

Then

- (i) $\mathcal{C}^c \equiv \{C^c : C \in \mathcal{C}\}$ is VC with $V(\mathcal{C}^c) = V(\mathcal{C})$;
- (ii) $\mathcal{C} \cap \mathcal{D} \equiv \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC with index $\leq V_{\mathcal{C}} + V_{\mathcal{D}} - 1$;
- (iii) $\mathcal{C} \sqcup \mathcal{D} \equiv \{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC with index $\leq V_{\mathcal{C}} + V_{\mathcal{D}} - 1$;
- (iv) $\mathcal{D} \times \mathcal{E}$ is VC in $\mathcal{X} \times \mathcal{W}$ with VC index $\leq V_{\mathcal{D}} + V_{\mathcal{E}} - 1$;

(v) $\phi(\mathcal{C})$ is VC with index $V_{\mathcal{C}}$ if ϕ is one-to-one;

(vi) $\psi^{-1}(\mathcal{C})$ is VC with index $\leq V_{\mathcal{C}}$.

LEMMA 3. *For any class \mathcal{C} of sets in a set \mathcal{X} , the class $\mathcal{F}_{\mathcal{C}}$ of indicator functions of sets in \mathcal{C} is VC-subgraph if and only if \mathcal{C} is a VC-class.*

Moreover, whenever at least one of \mathcal{C} or $\mathcal{F}_{\mathcal{C}}$ is VC, the respective VC-indices are equal.

Proof. Let \mathcal{D} be the collection of sets of the form

$$\{(x, t) : t < \mathbf{1}\{x \in C\}\}$$

for all $C \in \mathcal{C}$.

Suppose that \mathcal{D} is VC, and let $k = V(\mathcal{D})$.

Then no set of the form $\{(x_1, 0), \dots, (x_k, 0)\}$ can be shattered by \mathcal{D} , and hence $V(\mathcal{C}) \leq V(\mathcal{D})$.

Now suppose that \mathcal{C} is VC with VC-index k .

Since for any $t < 0$, $\mathbf{1}\{x \in C\} > t$ for all x and all C , we have that no collection

$$\{(x_1, t_1), \dots, (x_k, t_k)\}$$

can be shattered by \mathcal{D} if any of the t_j s are < 0 .

It is similarly true that no collection

$$\{(x_1, t_1), \dots, (x_k, t_k)\}$$

can be shattered by \mathcal{D} if any of the t_j s are ≥ 1 , since $\mathbf{1}\{x \in C\} > t$ is never true when $t \geq 1$.

It can now be deduced that

$$\{(x_1, t_1), \dots, (x_k, t_k)\}$$

can only be shattered if

$$\{(x_1, 0), \dots, (x_k, 0)\}$$

can be shattered.

But this can only happen if $\{x_1, \dots, x_k\}$ can be shattered by \mathcal{C} .

Thus $V(\mathcal{D}) \leq V(\mathcal{C})$. \square

LEMMA 4. Let \mathcal{F} and \mathcal{G} be VC-subgraph classes of functions on a set \mathcal{X} , with respective VC indices $V_{\mathcal{F}}$ and $V_{\mathcal{G}}$.

Let $g : \mathcal{X} \mapsto \mathbb{R}$, $\phi : \mathbb{R} \mapsto \mathbb{R}$, and $\psi : \mathcal{Z} \mapsto \mathcal{X}$ be fixed functions.

Then

- (i) $\mathcal{F} \wedge \mathcal{G} \equiv \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is VC-subgraph with index $\leq V_{\mathcal{F}} + V_{\mathcal{G}} - 1$;
- (ii) $\mathcal{F} \vee \mathcal{G}$ is VC with index $\leq V_{\mathcal{F}} + V_{\mathcal{G}} - 1$;
- (iii) $\{\mathcal{F} > 0\} \equiv \{\{f > 0\} : f \in \mathcal{F}\}$ is a VC-class of sets with index $V_{\mathcal{F}}$;
- (iv) $-\mathcal{F}$ is VC-subgraph with index $V_{\mathcal{F}}$;

- (v) $\mathcal{F} + g \equiv \{f + g : f \in \mathcal{F}\}$ is VC with index $V_{\mathcal{F}}$;
- (vi) $\mathcal{F} \cdot g \equiv \{fg : f \in \mathcal{F}\}$ is VC with index $\leq 2V_{\mathcal{F}} - 1$;
- (vii) $\mathcal{F} \circ \psi \equiv \{f(\psi) : f \in \mathcal{F}\}$ is VC with index $\leq V_{\mathcal{F}}$;
- (viii) $\phi \circ \mathcal{F}$ is VC with index $\leq V_{\mathcal{F}}$ for monotone ϕ .

The next two lemmas refer to properties of monotone processes and classes of monotone functions.

LEMMA 5. *Let $\{X(t), t \in T\}$ be a monotone increasing stochastic process, where $T \subset \mathbb{R}$.*

Then X is VC-subgraph with index $V(X) = 2$.

Proof. Let \mathcal{X} be the set of all monotone increasing functions $g : T \mapsto \mathbb{R}$; and for any $s \in T$ and $x \in \mathcal{X}$, define $(s, x) \mapsto f_s(x) = x(s)$.

Thus the proof is complete if we can show that the class of functions $\mathcal{F} \equiv \{f_s : s \in T\}$ is VC-subgraph with VC index 2.

Now let $(x_1, t_1), (x_2, t_2)$ be any two points in $\mathcal{X} \times \mathbb{R}$.

\mathcal{F} shatters $(x_1, t_1), (x_2, t_2)$ if the graph \mathcal{G} of $(f_s(x_1), f_s(x_2))$ in \mathbb{R}^2 “surrounds” the point (t_1, t_2) as s ranges over T .

By surrounding a point $(a, b) \in \mathbb{R}^2$, we mean that the graph must pass through all four of the sets $\{(u, v) : u \leq a, v \leq b\}$, $\{(u, v) : u > a, v \leq b\}$, $\{(u, v) : u \leq a, v > b\}$ and $\{(u, v) : u > a, v > b\}$.

By the assumed monotonicity of x_1 and x_2 , the graph \mathcal{G} forms a monotone curve in \mathbb{R}^2 , and it is thus impossible for it to surround any point in \mathbb{R}^2 .

Thus $(x_1, t_1), (x_2, t_2)$ cannot be shattered by \mathcal{F} , and the desired result follows. \square

LEMMA 6. *The set \mathcal{F} of all monotone functions $f : \mathbb{R} \mapsto [0, 1]$ satisfies*

$$\sup_Q \log N(\epsilon, \mathcal{F}, L_2(Q)) \leq \frac{K}{\epsilon}, \quad 0 < \epsilon < 1,$$

where the supremum is taken over all probability measures Q , and the constant $K < \infty$ is universal.

BUEI Classes

Recall for a class of measurable functions \mathcal{F} , with envelope F , the uniform entropy integral

$$J(\delta, \mathcal{F}, L_2) \equiv \int_0^\delta \sqrt{\sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon,$$

where the supremum is taken over all finitely discrete probability measures Q with $\|F\|_{Q,2} > 0$.

Note the dependence on choice of envelope F .

This is crucial since there are many random functions which can serve as an envelope.

For example, if F is an envelope, then so is $F + 1$ and $2F$.

One must allow that different envelopes may be needed in different settings.

We say that the class \mathcal{F} has *bounded uniform entropy integral* (BUEI) with envelope F —or is *BUEI* with envelope F —if $J(1, \mathcal{F}, L_2) < \infty$ for that particular choice of envelope.

Theorem 2 tells us that a VC-class \mathcal{F} is automatically BUEI with any envelope.

We leave it as an exercise to show that if \mathcal{F} and \mathcal{G} are BUEI with respective envelopes F and G , then $\mathcal{F} \sqcup \mathcal{G}$ is BUEI with envelope $F \vee G$.

LEMMA 7. Let $\mathcal{F}_1, \dots, \mathcal{F}_k$ be BUEI classes with respective envelopes F_1, \dots, F_k , and let $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ satisfy

$$|\phi \circ f(x) - \phi \circ g(x)|^2 \leq c^2 \sum_{j=1}^k (f_j(x) - g_j(x))^2, \quad (1)$$

for every $f, g \in \mathcal{F}_1 \times \dots \times \mathcal{F}_k$ and x for a constant $0 < c < \infty$.

Then the class $\phi \circ (\mathcal{F}_1, \dots, \mathcal{F}_k)$ is BUEI with envelope

$$H \equiv |\phi(f_0)| + c \sum_{j=1}^k (|f_{0j}| + F_j),$$

where $f_0 \equiv (f_{01}, \dots, f_{0k})$ is any function in $\mathcal{F}_1 \times \dots \times \mathcal{F}_k$, and where $\phi \circ (\mathcal{F}_1, \dots, \mathcal{F}_k)$ is as defined in Lemma 8.10.

Some useful consequences of Lemma 7 are given in the following lemma:

LEMMA 8. *Let \mathcal{F} and \mathcal{G} be BUEI with respective envelopes F and G , and let $\phi : \mathbb{R} \mapsto \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant $0 < c < \infty$.*

Then

- (i) $\mathcal{F} \wedge \mathcal{G}$ is BUEI with envelope $F + G$;*
- (ii) $\mathcal{F} \vee \mathcal{G}$ is BUEI with envelope $F + G$;*
- (iii) $\mathcal{F} + \mathcal{G}$ is BUEI with envelope $F + G$;*
- (iv) $\phi(\mathcal{F})$ is BUEI with envelope $|\phi(f_0)| + c(|f_0| + F)$, provided $f_0 \in \mathcal{F}$.*

Most of the BUEI preservation results we give in this section have parallel Donsker preservation properties.

An important exception, and one which is perhaps the primary justification for the use of BUEI preservation techniques, applies to products of Donsker classes.

As verified in the following theorem, the product of two BUEI classes is BUEI, whether or not the two classes involved are bounded:

THEOREM 4. *Let \mathcal{F} and \mathcal{G} be BUEI classes with respective envelopes F and G .*

Then

$$\mathcal{F} \cdot \mathcal{G} \equiv \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$$

is BUEI with envelope FG .

In order for BUEI results to be useful for obtaining Donsker results, it is necessary that sufficient measurability be established so that Theorem 8.19 can be used.

Pointwise measurability (PM) is sufficient measurability for this purpose.

Since there are significant similarities between PM preservation and BUEI preservation results, one can construct useful joint PM and BUEI preservation results:

LEMMA 9. *Let the classes $\mathcal{F}_1, \dots, \mathcal{F}_k$ be both BUEI and PM with respective envelopes F_1, \dots, F_k , and let $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ satisfy (1) for every*

$$f, g \in \mathcal{F}_1 \times \dots \times \mathcal{F}_k$$

and x for a constant $0 < c < \infty$.

Then the class $\phi \circ (\mathcal{F}_1, \dots, \mathcal{F}_k)$ is both BUEI and PM with envelope

$$H \equiv |\phi(f_0)| + c \sum_{j=1}^k (|f_{0j}| + F_j),$$

where f_0 is any function in $\mathcal{F}_1 \times \dots \times \mathcal{F}_k$.

LEMMA 10. *Let the classes \mathcal{F} and \mathcal{G} be both BUEI and PM with respective envelopes F and G , and let $\phi : \mathbb{R} \mapsto \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant $0 < c < \infty$.*

Then

- (i) $\mathcal{F} \cup \mathcal{G}$ is both BUEI and PM with envelope $F \vee G$;*
- (ii) $\mathcal{F} \wedge \mathcal{G}$ is both BUEI and PM with envelope $F + G$;*
- (iii) $\mathcal{F} \vee \mathcal{G}$ is both BUEI and PM with envelope $F + G$;*
- (iv) $\mathcal{F} + \mathcal{G}$ is both BUEI and PM with envelope $F + G$;*
- (v) $\mathcal{F} \cdot \mathcal{G}$ is both BUEI and PM with envelope FG ;*
- (vi) $\phi(\mathcal{F})$ is both BUEI and PM with envelope $|\phi(f_0)| + c(|f_0| + F)$,
where $f_0 \in \mathcal{F}$.*

If a class of measurable functions \mathcal{F} is both BUEI and PM with envelope F , then Theorem 8.19 implies that \mathcal{F} is P -Donsker whenever $P^* F^2 < \infty$.

Note that we have somehow avoided discussing preservation for subsets of classes.

This is because it is unclear whether a subset of a PM class \mathcal{F} is itself a PM class.

The difficulty is that while \mathcal{F} may have a countable dense subset \mathcal{G} (dense in terms of pointwise convergence), it is unclear whether any arbitrary subset $\mathcal{H} \subset \mathcal{F}$ also has a suitable countable dense subset.

An easy way around this problem is to use various preservation results to establish that \mathcal{F} is P -Donsker, and then it follows directly that any $\mathcal{H} \subset \mathcal{F}$ is also P -Donsker by the definition of weak convergence.

Bracketing Entropy

We now present several useful bracketing entropy results for certain function classes as well as a few preservation results.

We first mention that bracketing numbers are in general larger than covering numbers, as verified in the following lemma:

LEMMA 11. *Let \mathcal{F} be any class of real function on \mathcal{X} and $\|\cdot\|$ any norm on \mathcal{F} .*

Then

$$N(\epsilon, \mathcal{F}, \|\cdot\|) \leq N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$$

for all $\epsilon > 0$.

Proof. Fix $\epsilon > 0$, and let \mathcal{B} be collection of ϵ -brackets that covers \mathcal{F} .

From each bracket $B \in \mathcal{B}$, take a function $g(B) \in B \cap \mathcal{F}$ to form a finite collection of functions $\mathcal{G} \subset \mathcal{F}$ of the same cardinality as \mathcal{B} consisting of one function from each bracket in \mathcal{B} .

Now every $f \in \mathcal{F}$ lies in a bracket $B \in \mathcal{B}$ such that $\|f - g(B)\| \leq \epsilon$ by the definition of an ϵ -bracket.

Thus \mathcal{G} is an ϵ cover of \mathcal{F} of the same cardinality as \mathcal{B} . The desired conclusion now follows. \square

The first substantive bracketing entropy result we present considers classes of smooth functions on a bounded set $\mathcal{X} \subset \mathbb{R}^d$.

For any vector $k = (k_1, \dots, k_d)$ of nonnegative integers define the differential operator

$$D^k \equiv \frac{\partial^{|k|}}{(\partial x_1^{k_1}, \dots, \partial x_d^{k_d})},$$

where $|k| \equiv k_1 + \dots + k_d$.

As defined previously, let $\lfloor x \rfloor$ be the largest integer $j \leq x$, for any $x \in \mathbb{R}$.

For any function $f : \mathcal{X} \mapsto \mathbb{R}$ and $\alpha > 0$, define the norm

$$\|f\|_\alpha \equiv \max_{k:|k|\leq[\alpha]} \sup_x |D^k f(x)| + \max_{k:|k|=[\alpha]} \sup_{x,y} \frac{|D^k f(x) - D^k f(y)|}{\|x - y\|^{\alpha - [\alpha]}},$$

where the suprema are taken over $x \neq y$ in the interior of \mathcal{X} .

When $k = 0$, we set $D^k f = f$.

Now let $C_M^\alpha(\mathcal{X})$ be the set of all continuous functions $f : \mathcal{X} \mapsto \mathbb{R}$ with $\|f\|_\alpha \leq M$.

Recall that for a set A in a metric space, $\text{diam } A = \sup_{x,y \in A} d(x,y)$.

THEOREM 5. *Let $\mathcal{X} \subset \mathbb{R}^d$ be bounded and convex with nonempty interior.*

There exists a constant $K < \infty$ depending only on α , $\text{diam } \mathcal{X}$, and d such that

$$\log N_{[]}(\epsilon, C_1^\alpha(\mathcal{X}), L_r(Q)) \leq K \left(\frac{1}{\epsilon} \right)^{d/\alpha},$$

for every $r \geq 1$, $\epsilon > 0$, and any probability measure Q on \mathbb{R}^d .

We now consider several results for Lipschitz and Sobolev function classes.

We first present the results for covering numbers based on the uniform norm and then present the relationship to bracketing entropy:

THEOREM 6. *For a compact, convex subset $C \subset \mathbb{R}^d$, let \mathcal{F} be the class of all convex functions $f : C \mapsto [0, 1]$ with $|f(x) - f(y)| \leq L\|x - y\|$ for every x, y .*

For some integer $m \geq 1$, let \mathcal{G} be the class of all functions $g : [0, 1] \mapsto [0, 1]$ with $\int_0^1 [g^{(m)}(x)]^2 dx \leq 1$, where superscript (m) denotes the m 'th derivative.

Then

$$\log N(\epsilon, \mathcal{F}, \|\cdot\|_\infty) \leq K(1+L)^{d/2} \left(\frac{1}{\epsilon}\right)^{d/2}, \text{ and}$$

$$\log N(\epsilon, \mathcal{G}, \|\cdot\|_\infty) \leq M \left(\frac{1}{\epsilon}\right)^{1/m},$$

where $\|\cdot\|_\infty$ is the uniform norm and the constant $K < \infty$ depends only on d and C and the constant M depends only on m .

The following lemma shows how Theorem 6 applies to bracketing entropy:

LEMMA 12. *For any norm $\| \cdot \|$ dominated by $\| \cdot \|_\infty$ and any class of functions \mathcal{F} ,*

$$\log N_{[]} (2\epsilon, \mathcal{F}, \| \cdot \|) \leq \log N(\epsilon, \mathcal{F}, \| \cdot \|_\infty),$$

for all $\epsilon > 0$.

Proof. Let f_1, \dots, f_m be a uniform ϵ -cover of \mathcal{F} .

Since the 2ϵ -brackets $[f_i - \epsilon, f_i + \epsilon]$ now cover \mathcal{F} , the result follows. \square

We now present a second Lipschitz continuity result which is in fact a generalization of Lemma 12.

The result applies to function classes of the form $\mathcal{F} = \{f_t : t \in T\}$, where

$$|f_s(x) - f_t(x)| \leq d(s, t)F(x) \quad (2)$$

for some metric d on T , some real function F on the sample space \mathcal{X} , and for all $x \in \mathcal{X}$.

This special Lipschitz structure arises in a number of settings, including parametric Z- and M- estimation.

For example, consider the least absolute deviation regression setting of Section 2.2.6, under the assumption that the random covariate U and regression parameter θ are constrained to known compact subsets $\mathcal{U}, \Theta \subset \mathbb{R}^p$.

Recall that, in this setting, the outcome given U is modeled as $Y = \theta'U + e$, where the residual error e has median zero.

Estimation of the true parameter value θ_0 is accomplished by minimizing $\theta \mapsto \mathbb{P}_n m_\theta$, where

$$m_\theta(X) \equiv |e - (\theta - \theta_0)'U| - |e|, \quad X \equiv (Y, U)$$

and $e = Y - \theta_0'U$.

From (2.20) in Section 2.2.6, we know that the class $\mathcal{F} = \{m_\theta : \theta \in \Theta\}$ satisfies (2) with $T = \Theta$, $d(s, t) = \|s - t\|$ and $F(x) = \|u\|$, where $x = (y, u)$ is a realization of X .

The following theorem shows that the bracketing numbers for a general \mathcal{F} satisfying (2) are bounded by the covering numbers for the associated index set T :

THEOREM 7. *Suppose the class of functions $\mathcal{F} = \{f_t : t \in T\}$ satisfies (2) for every $s, t \in T$ and some fixed function F .*

Then, for any norm $\|\cdot\|$,

$$N_{[]} (2\epsilon \|F\|, \mathcal{F}, \|\cdot\|) \leq N(\epsilon, T, d).$$

Proof. Note that for any ϵ -net t_1, \dots, t_k that covers T with respect to d , the brackets $[f_{t_j} - \epsilon F, f_{t_j} + \epsilon F]$ cover \mathcal{F} .

Since these brackets are all of size $2\epsilon\|F\|$, the proof is complete. \square

Note that when $\|\cdot\|$ is any norm dominated by $\|\cdot\|_\infty$, Theorem 7 simplifies to Lemma 12 when $T = \mathcal{F}$ and $d = \|\cdot\|_\infty$ (and thus automatically $F = 1$).

We move now from continuous functions to monotone functions:

THEOREM 8. *For each integer $r \geq 1$, there exists a constant $K < \infty$ such that the class \mathcal{F} of monotone functions $f : \mathbb{R} \mapsto [0, 1]$ satisfies*

$$\log N_{[]}(\epsilon, \mathcal{F}, L_r(Q)) \leq \frac{K}{\epsilon},$$

for all $\epsilon > 0$ and every probability measure Q .

Useful preservation results for bracketing entropy are, unfortunately, rare, but the following are two such results:

LEMMA 13. *Let \mathcal{F} and \mathcal{G} be classes of measurable function.*

Then for any probability measure Q and any $1 \leq r \leq \infty$,

(i) $N_{[]} (2\epsilon, \mathcal{F} + \mathcal{G}, L_r(Q)) \leq N_{[]} (\epsilon, \mathcal{F}, L_r(Q)) N_{[]} (\epsilon, \mathcal{G}, L_r(Q));$

(ii) *Provided \mathcal{F} and \mathcal{G} are bounded by 1,*

$$N_{[]} (2\epsilon, \mathcal{F} \cdot \mathcal{G}, L_r(Q)) \leq N_{[]} (\epsilon, \mathcal{F}, L_r(Q)) N_{[]} (\epsilon, \mathcal{G}, L_r(Q)).$$