Introduction to Empirical Processes and Semiparametric Inference Lecture 12: Glivenko-Cantelli and Donsker Results

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Glivenko-Cantelli Results

The existence of an integrable envelope of the centered functions of a class $\mathcal{F}$ is a necessary condition for $\mathcal{F}$ to be P-G-C:

LEMMA 1. *If the class of functions $\mathcal{F}$ is strong P-G-C, then $P\|f - Pf\|^*_{\mathcal{F}} < \infty$.*

If in addition $\|P\|_{\mathcal{F}} < \infty$, then also $P\|f\|^*_{\mathcal{F}} < \infty$.

THEOREM 1. *Let $\mathcal{F}$ be a class of measurable functions and suppose that*

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$$

for every $\epsilon > 0$.

Then $\mathcal{F}$ is P-Glivenko-Cantelli.

Proof. Fix $\epsilon > 0$.

Since the L_1-bracketing entropy is bounded, it is possible to choose finitely many ϵ-brackets $[l_i, u_i]$ so that

- their union contains $\mathcal{F}$
- and $P(u_i - l_i) < \epsilon$ for every i.

Now, for every $f \in \mathcal{F}$, there is a bracket $[l_i, u_i]$ containing f with

$$(\mathbb{P}_n - P)f \leq (\mathbb{P}_n - P)u_i + P(u_i - f) \leq (\mathbb{P}_n - P)u_i + \epsilon.$$

Hence

$$\begin{aligned}\sup_{f\in\mathcal{F}}(\mathbb{P}_n - P)f &\leq \max_i(\mathbb{P}_n - P)u_i + \epsilon \\ &\stackrel{\text{as}*}{\longrightarrow} \epsilon.\end{aligned}$$

Similar arguments can be used to verify that

$$\begin{aligned}\inf_{f\in\mathcal{F}}(\mathbb{P}_n - P)f &\geq \min_i(\mathbb{P}_n - P)l_i - \epsilon \\ &\stackrel{\text{as}*}{\longrightarrow} -\epsilon.\end{aligned}$$

The desired result now follows since ϵ was arbitrary.$\square$

THEOREM 2. *Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope F and*

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

for every $\epsilon > 0$, where the supremum is taken over all finite probability measures Q with $\|F\|_{Q,1} > 0$.

If $P^* F < \infty$, then $\mathcal{F}$ is P-G-C.

Proof. The result is trivial if $P^*F = 0$.

Hence we will assume without loss of generality that $P^*F > 0$.

Thus there exists an $\eta > 0$ such that, with probability 1, $\mathbb{P}_n F > \eta$ for all n large enough.

Fix $\epsilon > 0$.

By assumption, there is a $K < \infty$ such that

$$\mathbf{1}\{\mathbb{P}_n F > 0\} \log N(\epsilon \mathbb{P}_n F, \mathcal{F}, L_1(\mathbb{P}_n)) \leq K$$

almost surely, since $\mathbb{P}_n$ is a finite probability measure.

Hence, with probability 1,

$$\log N(\epsilon\eta, \mathcal{F}, L_1(\mathbb{P}_n)) \leq K$$

for all n large enough.

Since ϵ was arbitrary, we now have that

$$\log N(\epsilon, \mathcal{F}, L_1(\mathbb{P}_n)) = O_P^*(1)$$

for all $\epsilon > 0$.

Now fix $\epsilon > 0$ (again) and $M < \infty$, and define

$$\mathcal{F}_M \equiv \{f \mathbf{1}\{F \leq M\} : f \in \mathcal{F}\}.$$

Since,

$$\|(f - g)\mathbf{1}\{F \leq M\}\|_{1,\mathbb{P}_n} \leq \|f - g\|_{1,\mathbb{P}_n}$$

for any $f, g \in \mathcal{F}$, we have

$$N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) \leq N(\epsilon, \mathcal{F}, L_1(\mathbb{P}_n)).$$

Hence

$$\log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = O_P^*(1).$$

Finally, since ϵ and M are both arbitrary, the desired result follows from Theorem 3 below.$\square$

THEOREM 3. *Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope F such that $P^*F < \infty$.*

Let $\mathcal{F}_M$ be as defined above.

If

$$\log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_P^*(n)$$

for every $\epsilon > 0$ and $M < \infty$, then $P\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \to 0$ and $\mathcal{F}$ is strong P-G-C.

Donsker Results

The following lemma outlines several properties of Donsker classes:

LEMMA 2. *Let $\mathcal{F}$ be a class of measurable functions, with envelope $F \equiv \|f\|_{\mathcal{F}}$.*

For any $f, g \in \mathcal{F}$, define $\rho(f, g) \equiv \left\{P(f - Pf - g + Pg)^2\right\}^{1/2}$; and, for any $\delta > 0$, let $\mathcal{F}_\delta \equiv \{f - g : \rho(f, g) < \delta\}$.

Then the following are equivalent:

(i) $\mathcal{F}$ is P-Donsker;

(ii) $(\mathcal{F}, \rho)$ *is totally bounded and* $\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \xrightarrow{\mathsf{P}} 0$ *for every* $\delta_n \downarrow 0$*;*

(iii) $(\mathcal{F}, \rho)$ *is totally bounded and* $\mathsf{E}^*\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \to 0$ *for every* $\delta_n \downarrow 0$*.*

These conditions imply that

- $\mathsf{E}^*\|\mathbb{G}_n\|_{\mathcal{F}}^r \to \mathsf{E}\|\mathbb{G}\|_{\mathcal{F}}^r < \infty$, *for every* $0 < r < 2$*;*
- $P\left(\|f - Pf\|_{\mathcal{F}}^* > x\right) = o(x^{-2})$ *as* $x \to \infty$*; and*
- $\mathcal{F}$ *is strong* P*-G-C.*

If in addition $\|P\|_{\mathcal{F}} < \infty$, *then also* $P(F^* > x) = o(x^{-2})$ *as* $x \to \infty$*.*

Recall the following bracketing entropy Donsker theorem from Chapter 2:

THEOREM 4. *Let $\mathcal{F}$ be a class of measurable functions with*

$$J_{[]}(\infty, \mathcal{F}, L_2(P)) < \infty.$$

Then $\mathcal{F}$ is P-Donsker.

We omit the proof.

The following is the correctly stated uniform entropy Donsker theorem:

THEOREM 5. *Let $\mathcal{F}$ be a class of measurable functions with envelope F and $J(1, \mathcal{F}, L_2) < \infty$.*

Let the classes $\mathcal{F}_\delta$ and $\mathcal{F}_\infty^2 \equiv \{h^2 : h \in \mathcal{F}_\infty\}$ be P-measurable for every $\delta > 0$.

If $P^* F^2 < \infty$, then $\mathcal{F}$ is P-Donsker.

We note here that by Proposition 8.11, if $\mathcal{F}$ is PM, then so are

- $\mathcal{F}_\delta$ and
- $\mathcal{F}_\infty^2$,

for all $\delta > 0$, provided $\mathcal{F}$ has envelope F such that $P^* F^2 < \infty$.

Since PM implies P-measurability, all measurability requirements for Theorem 5 are thus satisfied whenever $\mathcal{F}$ is PM.

Proof of Theorem 5. Let the positive, decreasing sequence $\delta_n \downarrow 0$ be arbitrary.

By Markov's inequality for outer probability (see Lemma 6.10) and the symmetrization Theorem 8.8,

$$\mathrm{P}^* \left(\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} > x \right) \leq \frac{2}{x} \mathsf{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i f(X_i) \right\|_{\mathcal{F}_{\delta_n}},$$

for i.i.d. Rademachers $\epsilon_1, \ldots, \epsilon_n$ independent of $X_1, \ldots, X_n$.

By the P-measurability assumption for $\mathcal{F}_\delta$, for all $\delta > 0$,

- the standard version of Fubini's theorem applies, and
- the outer expectation is just a standard expectation and can be calculated in the order $\mathrm{E}_X \mathrm{E}_\epsilon$.

Accordingly, fix $X_1, \ldots, X_n$.

By Hoeffding's inequality (Lemma 8.7), the stochastic process $f \mapsto n^{-1/2} \times \sum_{i=1}^n \epsilon_i f(X_i)$ is sub-Gaussian for the $L_2(\mathbb{P}_n)$-seminorm

$$\|f\|_n \equiv \sqrt{\frac{1}{n}\sum_{i=1}^n f^2(X_i)}.$$

This stochastic process is also separable since, for any measure Q and $\epsilon > 0$,

$$\begin{aligned} N(\epsilon, \mathcal{F}_{\delta_n}, L_2(Q)) &\leq N(\epsilon, \mathcal{F}_\infty, L_2(Q)) \\ &\leq N^2(\epsilon/2, \mathcal{F}, L_2(Q)), \end{aligned}$$

and the latter is finite for any finite dimensional probability measure Q and any $\epsilon > 0$.

Thus the second conclusion of Corollary 8.5 holds with

$$\mathsf{E}_\epsilon \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i f(X_i) \right\|_{\mathcal{F}_{\delta_n}} \lesssim \int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}_{\delta_n}, L_2(\mathbb{P}_n))} d\epsilon. \quad (1)$$

Note that we can omit the term

$$\mathsf{E}\left|n^{-1/2}\sum_{i=1}^{n}\epsilon_i f_0(X_i)\right|$$

from the conclusion of the corollary because $0 = f - f \in \mathcal{F}_{\delta_n}$.

For sufficiently large ϵ, the set $\mathcal{F}_{\delta_n}$ fits into a single ball of $L_2(\mathbb{P}_n)$-radius ϵ around the origin, in which case the integrand on the right-hand-side of (1) is zero.

This will definitely happen when ϵ is larger than θ_n, where

$$\theta_n^2 \equiv \sup_{f\in\mathcal{F}_{\delta_n}} \|f\|_n^2 = \left\|\frac{1}{n}\sum_{i=1}^{n} f^2(X_i)\right\|_{\mathcal{F}_{\delta_n}}.$$

Thus the right-hand-side of (1) is bounded by

$$
\begin{aligned}
&\int_0^{\theta_n} \sqrt{\log N(\epsilon, \mathcal{F}_{\delta_n}, L_2(\mathbb{P}_n))}d\epsilon \\
&\quad\lesssim \int_0^{\theta_n} \sqrt{\log N^2(\epsilon/2, \mathcal{F}, L_2(\mathbb{P}_n))}d\epsilon \\
&\quad\lesssim \int_0^{\theta_n/(2\|F\|_n)} \sqrt{\log N(\epsilon\|F\|_n, \mathcal{F}, L_2(\mathbb{P}_n))}d\epsilon\|F\|_n \\
&\quad\lesssim \|F\|_n J(\theta_n, \mathcal{F}, L_2).
\end{aligned}
$$

The second inequality follows from the change of variables $u = \epsilon/(2\|F\|_n)$ (and then renaming u to ϵ).

For the third inequality, note that we can add $1/2$ to the envelope function F without changing the existence of its second moment.

Hence $\|F\|_n \geq 1/2$ without loss of generality, and thus

$$\theta_n/(2\|F\|_n) \leq \theta_n.$$

Because $\|F\|_n = O_p(1)$, we can now conclude that the left-hand-side of (1) goes to zero in probability, provided we can verify that $\theta_n \xrightarrow{\mathsf{P}} 0$.

This would then imply asymptotic $L_2(P)$-equicontinuity in probability.

Since $\|Pf^2\|_{\mathcal{F}_{\delta_n}} \to 0$ and $\mathcal{F}_{\delta_n} \subset \mathcal{F}_\infty$, establishing that

$$\|\mathbb{P}_n - P\|_{\mathcal{F}_\infty^2} \xrightarrow{\mathsf{P}} 0$$

would prove that $\theta_n \xrightarrow{\mathsf{P}} 0$.

The class $\mathcal{F}_\infty^2$ has integrable envelope $(2F)^2$ and is P-measurable by assumption.

Since also, for any $f, g \in \mathcal{F}_\infty$,

$$\mathbb{P}_n|f^2 - g^2| \leq \mathbb{P}_n(|f-g|4F) \leq \|f-g\|_n\|4F\|_n,$$

we have that the covering number

$$N(\epsilon\|2F\|_n^2, \mathcal{F}_\infty^2, L_1(\mathbb{P}_n))$$

is bounded by

$$N(\epsilon\|F\|_n, \mathcal{F}_\infty, L_2(\mathbb{P}_n)).$$

Since this last covering number is bounded by

$$\sup_Q N^2(\epsilon\|F\|_{Q,2}/2, \mathcal{F}, L_2(Q)) < \infty,$$

where the supremum is taken over all finitely discrete probability measures with $\|F\|_{Q,2} > 0$, we have by Theorem 2 that $\mathcal{F}_\infty^2$ is P-Glivenko-Cantelli.

Thus $\hat{\theta}_n \xrightarrow{\mathsf{P}} 0$.

This completes the proof of asymptotic equicontinuity.

The last thing we need to prove is that $\mathcal{F}$ is totally bounded in $L_2(P)$.

By the result of the last paragraph, there exists a sequence of discrete probability measures P_n with

$$\|(P_n - P)f^2\|_{\mathcal{F}_\infty} \to 0.$$

Fix $\epsilon > 0$ and take n large enough so that

$$\|(P_n - P)f^2\|_{\mathcal{F}_\infty} < \epsilon^2.$$

Note that

$$N(\epsilon, \mathcal{F}, L_2(P_n))$$

is finite by assumption, and, for any $f, g \in \mathcal{F}$ with $\|f - g\|_{P_n,2} < \epsilon$,

$$P(f-g)^2 \leq P_n(f-g)^2 + |(P_n - P)(f-g)^2| \leq 2\epsilon^2.$$

Thus any ϵ-net in $L_2(P_n)$ is also a $\sqrt{2}\epsilon$-net in $L_2(P)$.

Hence $\mathcal{F}$ is totally bounded in $L_2(P)$ since ϵ was arbitrary.$\square$

Entropy Calculations

The focus of this chapter is on computing entropy for empirical processes.

An important use of such entropy calculations is in evaluating whether a class of functions $\mathcal{F}$ is Glivenko-Cantelli and/or Donsker or neither.

Some of these uses will be very helpful in Chapter 14 for establishing rates of convergence for M-estimators.

We begin the chapter by describing methods to evaluate uniform entropy.

Provided the uniform entropy for a class $\mathcal{F}$ is not too large, $\mathcal{F}$ might be G-C or Donsker, as long as sufficient measurability holds.

One can think of this chapter as a handbag of tools for establishing weak convergence properties of empirical processes.

Vapnik-Červonenkis (VC) Classes

In this section, we introduce Vapnik-Červonenkis (VC) classes of sets, VC-classes of functions, and several related function classes.

We then present several examples of VC-classes.

Consider an arbitrary collection $\{x_1, \ldots, x_n\}$ of points in a set $\mathcal{X}$ and a collection $\mathcal{C}$ of subsets of $\mathcal{X}$.

We say that $\mathcal{C}$ *picks out* a certain subset A of $\{x_1, \ldots, x_n\}$ if $A = C \cap \{x_1, \ldots, x_n\}$ for some $C \in \mathcal{C}$.

We say that $\mathcal{C}$ *shatters* $\{x_1, \ldots, x_n\}$ if all of the 2^n possible subsets of $\{x_1, \ldots, x_n\}$ are picked out by the sets in $\mathcal{C}$.

The *VC-index* $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest n for which no set of size n $\{x_1, \ldots, x_n\} \subset \mathcal{X}$ is shattered by $\mathcal{C}$.

If $\mathcal{C}$ shatters all sets $\{x_1, \ldots, x_n\}$ for all $n \geq 1$, we set $V(\mathcal{C}) = \infty$.

Clearly, the more refined $\mathcal{C}$ is, the higher the VC-index.

We say that $\mathcal{C}$ is a *VC-class* if $V(\mathcal{C}) < \infty$.

For example, let $\mathcal{X} = \mathbb{R}$ and define the collection of sets $\mathcal{C} = \{(-\infty, c] : c \in \mathbb{R}\}$.

Consider any two point set $\{x_1, x_2\} \subset \mathbb{R}$ and assume, without loss of generality, that $x_1 < x_2$.

It is easy to verify that $\mathcal{C}$ can pick out the null set $\{\}$ and the sets $\{x_1\}$ and $\{x_1, x_2\}$ but can't pick out $\{x_2\}$.

Thus $V(\mathcal{C}) = 2$ and $\mathcal{C}$ is a VC-class.

As another example, let $\mathcal{C} = \{(a, b] : -\infty \leq a < b \leq \infty\}$.

The collection can shatter any two point set, but consider what happens with a three point set $\{x_1, x_2, x_3\}$.

Without loss of generality, assume $x_1 < x_2 < x_3$, and note that the set $\{x_1, x_3\}$ cannot be picked out with $\mathcal{C}$.

Thus $V(\mathcal{C}) = 3$ in this instance.

For any class of sets $\mathcal{C}$ and any collection $\{x_1, \ldots, x_n\} \subset \mathcal{X}$, let $\Delta_n(\mathcal{C}, x_1, \ldots, x_n)$ be the number of subsets of $\{x_1, \ldots, x_n\}$ which can be picked out by $\mathcal{C}$.

A surprising combinatorial result is that if $V(\mathcal{C}) < \infty$, then $\Delta_n(\mathcal{C}, x_1, \ldots, x_n)$ can increase in n no faster than $O(n^{V(\mathcal{C})-1})$.

This is more precisely stated in the following lemma:

LEMMA 3. *For a VC-class of sets $\mathcal{C}$,*

$$\max_{x_1,\ldots,x_n \in \mathcal{X}} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) \leq \sum_{j=1}^{V(\mathcal{C})-1} \begin{pmatrix} n \\ j \end{pmatrix}.$$

Since the right-hand-side is bounded by $V(\mathcal{C})n^{V(\mathcal{C})-1}$, the left-hand-side grows polynomially of order at most $O(n^{V(\mathcal{C})-1})$.

This is a corollary of Lemma 2.6.2 of VW, and we omit the proof.

Let **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 **1**$\{\mathcal{C}\}$:

THEOREM 6. *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 7. *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 sconv$\mathcal{F}$).

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

The *convex hull* of a class of functions $\mathcal{F}$, denoted 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 conv$\mathcal{F}$ and $\overline{\text{sconv}}\mathcal{F}$ to denote the pointwise closure of 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 8. *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 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 8 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.*