A short note on the kernel VC-type condition

Yen-Chi Chen
University of Washington
September 1, 2020

This note is based on page 220-222 of the following book:

> [GN2016] Giné, E., & Nickl, R. (2016). Mathematical foundations of infinite-dimensional statistical models (Vol. 40). Cambridge University Press.

In the kernel density estimation (KDE), we observe IID random variables $X_1, \cdots, X_n$ from some unknown PDF f and we estimate the underlying PDF via

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right).$$

Note: We assume the dimension of the data is 1 to simplify the problem.

Let $f_h(x) = \mathbb{E}[\widehat{f}_h(x)]$ be its expectation and under mild conditions (e.g., 2-Hölder or bounded second derivatives), we have $f_h(x) = f(x) + O(h^2)$. Also, it is known in the literature that such estimator has uniform convergence

$$\sup_x |\widehat{f}_h(x) - f_h(x)| = O_P\left(\sqrt{\frac{|\log h|}{nh}}\right)$$

under some *kernel VC-type condition*. In this note, we will give a gentle discussion on sufficient conditions to the kernel VC conditions.

Let

$$\mathcal{K} = \left\{ y \mapsto K\left(\frac{x-y}{h}\right) : h > 0, x \in \mathbb{R} \right\}$$

be the collection of kernel functions indexed by x and h. For a collection of functions $\mathcal{F}$ and a metric of function ρ, we denote the ε-*covering number* of $\mathcal{F}$ under ρ as

$$N(\varepsilon, \mathcal{F}, \rho).$$

The ε-covering number is the least number of functions $f_1, \cdots, f_N$ such that any function $f \in \mathcal{F}$ will satisfy $\min_j \rho(f, f_j) \leq \varepsilon$. If such collection $f_1, \cdots, f_N$ attain this bound, we will call them an ε-cover of $\mathcal{F}$ under the metric ρ. Also, for a collection $\mathcal{F}$, *an envelope* F_0 of $\mathcal{F}$ is a function such that $f(x) \leq F_0(x)$ for all $f \in \mathcal{F}$.

Let $\|\cdot\|_{L_2(Q)}$ be the $L_2(Q)$ norm of functions, i.e.,

$$\|f\|_{L_2(Q)} = \int |f(x)|^2 dQ(x).$$

The kernel VC-type condition is the following condition:

(K) there exists an envelope F_0 and constants $A, v > 0$ such that

$$\sup_Q N(\varepsilon \|F_0\|_{L_2(Q)}, \mathcal{K}, \|\cdot\|_{L_2(Q)}) \leq \left(\frac{A}{\varepsilon}\right)^v. \tag{1}$$

This condition was first appear in the following seminal paper:

> [GG2002] Giné, E., & Guillou, A. (2002, November). Rates of strong uniform consistency for multivariate kernel density estimators. In Annales de l'Institut Henri Poincare (B) Probability and Statistics (Vol. 38, No. 6, pp. 907-921).

In [GG2002], the authors pointed out that if the kernel function $K(x)$ is of *bounded p-variation*, then the condition in equation (1) holds. A function f is of *bounded p-variation* if

$$v_p = \sup\left\{\sum_{j=1}^{n} |f(x_j - x_{j-1})|^p : -\infty < x_0 < x_1 < \cdots < x_n < \infty, n \in \mathbb{N}\right\}$$

is bounded. Most common kernel functions, such as Gaussian, Epanechnikov, cosine kernels are all of bounded p-variation. So it is a very mild condition.

Here we will give a high level idea on why bounded p-variation is enough to condition (K). Our explanation is based on Lemma 3.6.11 and Proposition 3.6.12 of [GN2016].

Lemma 1 (Lemma 3.6.11. of GN2016) *Let f be a function of bounded p-variation. Then there exists a non-decreasing functioin h such that $0 \leq h(x) \leq v_p(f)$ and a $1/p$-Hölder continuous function on the interval $[0, v_p(f)]$ such that $f = g \circ h$ and $\|g\|_\infty = \|f\|_\infty$.*

Note: a function f is called β-Hölder if there exists a constant L such that for any x, y, $|f(x) - f(y)| \leq L|x-y|^\beta$.

Proof (sketch).

We take $h(x)$ to be the *'vertical distance travelled'* of f until point x. Namely, let $I_z(x) = I(x \leq z)$. Then $h(x) = v_p(fI_x)$. By construction, h is non-decreasing and for any $x < y$, $|f(y) - f(x)|^p \leq h(y) - h(x)$ and $h(x) \in [0, v_p(f)]$.

Let $u \in [0, v_p(f)]$ be any possible value of h. Then we choose the function $g(u)$ to be the value of f that corresponds to any point in $h^{-1}(u)$. So by construction, $g \circ h(x) = g(h(x)) = f(x)$.

Now we verify that g is $1/p-$Hölder continuous. Consider $u, v \in [0, v_p(f)]$ such that $g(u) = f(x)$ and $g(v) = f(y)$. Then we have

$$|g(u) - g(v)| = |f(x) - f(y)| \leq |h(x) - h(y)|^{1/p} \leq |u - v|^{1/p}.$$

Thus, g is $1/p-$Hölder continuous.

□

Proposition 2 (Proposition 3.6.12. of GN2016 (simplified)) *Let f be a continuous function of bounded p-variation with $p \geq 1$. Consider the collection*

$$\mathcal{F} = \{x \mapsto f(tx - s) : t > 0, s \in \mathbb{R}\}.$$

Then $\mathcal{F}$ is of VC-type, i.e., there exists an envelop F_0 and positive numbers A, v such that for any probability measure Q,

$$N(\varepsilon \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q)) \leq \left(\frac{A}{\varepsilon}\right)^v.$$

With Proposition 2, one can easily see why bounded p-variation kernel implies the condition (K).

Proof of Proposition 2 (sketch).

By Lemma 1, we can write $f = g \circ h$, where h is non-decreasing and g is $1/p$-Hölder. Thus, any function in $\mathcal{F}$, $f(tx - s) = g(h(tx - s))$. We first consider the class

$$\mathcal{H} = \{x \mapsto h(tx - s) : t > 0, s \in \mathbb{R}\}.$$

Then $\mathcal{F}$ is just $1/p$-Hölder transform from $\mathcal{H}$.

Since f is continuous, h will also be continuous. Because of the non-decreasing property of h, we can define its generalized inverse $h^{-1}(u)$ for any value $u \in [0, v_p(f)]$.

The subgraph of a particular element indexed by t, s in $\mathcal{H}$ will be

$$\begin{aligned} G_{t,s} = \{(x,u) \in \mathbb{R} \times [0, v_p(f)] : u \leq h(tx - s)\} &= \{(x,u) \in \mathbb{R} \times [0, v_p(f)] : h^{-1}(u) \leq tx - s\} \\ &= \{(x,u) \in \mathbb{R} \times [0, v_p(f)] : h^{-1}(u) - tx + s \leq 0\}. \end{aligned}$$

Thus, all possible subgraphs in $\mathcal{H}$ is the set $\mathcal{G} = \{G_{t,s} : t > 0, s \in \mathbb{R}\}$. So the VC dimension of $\mathcal{H}$ is the VC dimension of the set $\mathcal{G}$.

Because each element $G_{t,s}$ is determined by the function $h^{-1}(u) - tx + s \leq 0$, one can easily see that

$$\begin{aligned} \mathcal{G} &\subset \mathcal{V} \\ \mathcal{V} &= \{\mathbb{V}_{a,b,c} : a, b, c \in \mathbb{R}\} \\ \mathbb{V}_{a,b,c} &= \{(x,u) : ah^{-1}(u) + bx + c \leq 0\}. \end{aligned}$$

Because $\mathcal{V}$ is formed by the vector space of 3 functions ($(x,u) \mapsto h^{-1}(u)$, $(x,u) \mapsto x$, $(x,u) \mapsto 1$), its VC dimension is at most 4 (see, e.g., Proposition 3.6.6. of [GN2016]). So the VC dimension of $\mathcal{G} \subset \mathcal{V}$ will be at most 4, which implies that $\mathcal{H}$ is a VC-type class with VC dimension at most 4 and we can pick the envelope function of $\mathcal{H}$ to be the constant $v_p(f)$.

Then by the Dudley-Pollard Theorem (see, e.g., Theorem 3.6.9 of [GN2016]), there exist positive numbers A_0 such that

$$N(\varepsilon v_p(f), \mathcal{H}, L_2(Q)) \leq \left(\frac{A_0}{\varepsilon}\right)^5$$

for any probability measure Q. Note that the constant 5 comes from the fact that the VC dimension of the underlying subgraph $\mathcal{G}$ is at most 4.

Using the fact that g is $1/p$-Hölder so for any u, v with $\|u-v\|_{L_2(Q)} \leq \tau$,

$$\begin{aligned}
\|g(u)-g(v)\|_{L_2(Q)} &= \left(\int (g(u)-g(v))^2 dQ\right)^{1/2} \\
&\leq \left(\int |u-v|^{2/p} dQ\right)^{1/2} \\
&\leq c_p \tau^{1/p}
\end{aligned}$$

for some constant c_p. Thus, any ε-cover of $\mathcal{H}$ induces an $c_p \cdot \varepsilon^{1/p}$-cover of $\mathcal{F} = g \circ \mathcal{H}$ so we have

$$N(\varepsilon^{1/p} \cdot c_p \cdot v_p(f), \mathcal{F}, L_2(Q)) \leq N(\varepsilon v_p(f), \mathcal{H}, L_2(Q)) \leq \left(\frac{A_0}{\varepsilon}\right)^5,$$

which implies

$$N(\varepsilon F_0, \mathcal{F}, L_2(Q)) \leq \left(\frac{A}{\varepsilon}\right)^{5p},$$

where $F_0 = c_p^p \cdot v_p^p(f)$ is a contant envelope and $A = A_0^{1/p}$.

So we have completes the proof.

Note that this is a very loose bound–it can be improved a lot by the formal proof of Proposition 3.6.12. of [GN2016].

□