Lecture 11

In this lecture we will explore construction of "Honest Adaptive Confidence sets" in the context of density models. That is we will assume that we have data $X_1, \dots, X_n$ iid from a density η w.r.t lebesgue measure on $[0,1]$. η will be assumed to belong to $\mathcal{P}(\beta, M, B)$, $\beta > 0$. We want to find confidence sets for η with coverage ~~prob~~ probability $1-\alpha$ uniformly over all $\mathcal{P}(\beta, M, B)$, $\beta > 0$ ~~and~~ and "diameter" as small as possible.

Indeed if smoothness β was known, then we can construct an estimator $\hat{\eta}$ of η s.t. $\|\hat{\eta} - \eta\|_2 = O_p(n^{-\beta/2\beta+1})$. Therefore for a sufficiently large $C(\alpha)$, the set $\{\eta : \|\hat{\eta} - \eta\|_2 \le C(\alpha) n^{-\frac{\beta}{2\beta+1}}\}$ is $100 \times (1-\alpha)\%$ confidence set of the data generating density which shrinks in L_2 norm at the rate $n^{-\beta/2\beta+1}$. This raises the following questions.

- Does there exist confidence sets which shrink at faster rates than the given one for any β?
- If not, does there exist construct confidence sets whose construction does not depend on the knowledge of β?

We will try to understand these ~~two~~ type questions by discussing some theory of honest adaptive confidence sets in a general framework.

Consider a sequence of observations $X^{(n)}$ on measurable spaces (Ω_n, A_n), $n \in \mathbb{N}$. The model for distribution of $Z^{(n)}$ ~~is~~ consists of probability laws $P_\theta^{(n)}$, $\theta \in \Theta$ for some index set Θ. For the purpose of this topic assume that Θ is the subset of a separable Hilbert space (i.e. has a countable o.n.b.)

We wish to construct (asymptotic) confidence sets $\hat{C}_n$ of small diameter for the parameter θ with the following honesty property.

Honesty: $\{\hat{C}_n\}_{n \geq 1}$ is honest over Θ for a given confidence level $1-\alpha$ if

$$\liminf_{n\to\infty} \inf_{\theta \in \Theta} P_\theta(\theta \in \hat{C}_n) \geq 1-\alpha$$

We would like to construct confidence sets with the property

(i) The confidence set is honest of Θ.

(ii) The confidence set is centered at an estimator of choice $\hat{\theta}_m$.

(iii) The diameter of the confidence set adapts to submodels of Θ in a rate optimal way.

e.g. The set $\{\eta : \|\hat{\eta} - \eta\|_2 \leq c(\alpha) n^{-\frac{\beta}{2\beta+1}}\}$ does not have ~~a radi~~ has diameter $O(n^{-\beta/2\beta+1}) \gg n^{-\frac{\beta'}{2\beta'+1}}$ for any $\beta' > \beta$. So if the actual truth was smoother we are not adapting using this confidence set (even if $\hat{\eta}$ was an adaptive estimator.

The main result that we shall prove in this general setup can be described as follows. For a given submodel $\Theta_1 \subset \Theta$, the diameter of a confidence region that is honest over Θ cannot be of a smaller order, uniformly over Θ_1, than

(a) the "slowest rate" $\varepsilon_n \to 0$ such that for any estimator sequence $T_n(X^{(n)})$ and some $\alpha' > \alpha$

$$\liminf_{n\to\infty} \sup_{\theta \in \Theta_1} P_\theta\left(\|T_n - \theta\| \geq \varepsilon_n\right) > \alpha' \quad (\|\cdot\| \text{ is the norm on } \Theta)$$

This is typically the minimax rate of estimation for the model Θ_1.

(b) the minimax rate of testing (in the metric induced by $\|\cdot\|$) for the hypothesis

$H_0: \theta \in \Theta_1'$ vs $H_1: \theta \in \Theta: \|\theta - \Theta_1'\| > \varepsilon_n$ for any $\Theta_1' \subset \Theta$ (for example $\Theta_1' = \{\theta_1\}$ for $\theta_1 \in \Theta_1$. This rate is often governed by Θ instead of submodels Θ_1 of Θ)

e.g. Consider the density model and the problem of honest adaptive ~~est~~ confidence sets in L_2 over $\Theta = \mathcal{P}(\beta^*, M, B)$ for some given β^*. Let $\Theta_1 = \mathcal{P}(\beta_1, M, B)$ with $\beta_1 \geq \beta^*$. Then indeed $\Theta_1 \subseteq \Theta$. The above discussion implies that if a confidence set is honest over $\mathcal{P}(\beta^*, M, B)$, then its radius must be larger than the maximum of the following (over β_1)

(a) Minimax rate of estimation over $\mathcal{P}(\beta_1, M, B)$: $n^{-\frac{\beta_1}{2\beta_1+1}}$.

(b) Minimax rate of testing between

$H_0: \{\theta \in \mathcal{P}(\beta_1, M, B)\}$ vs $H_1: \{\theta \in \mathcal{P}(\beta^*, M, B): \|\theta - \theta'\|_2 \geq \varepsilon_n \ \forall \theta' \in \mathcal{P}(\beta_1, M, B)\}$

This turns out to be $\cancel{n^{-\frac{2\beta^*}{4\beta_1+1}}}$ $n^{-\frac{2\beta^*}{4\beta^*+1}}$

If $\beta_1 < 2\beta^*$ then (a) dominates and (b) dominates otherwise. Therefore for $\beta_1 > 2\beta^*$ one can no longer hope to have honest confidence sets over $\mathcal{P}(\beta^*, M, B)$ which shrinks at the rate of estimation rate over $\mathcal{P}(\beta_1, M, B)$. Before we delve into the details of the density model let us make the claims (a) & (b) in the general set up more formal.

(a) Lower Bound from Estimation Argument

Let $0<\alpha<\alpha'<1$ be given. Let $\{\varepsilon_n\}$ be a sequence of positive real numbers such that for every estimator sequence $\{T_n\}$,

$$\liminf_{n\to\infty} \sup_{\theta\in\Theta_1} P_\theta^{(n)}\left(\|T_n-\theta\| \geq \varepsilon_n\right) > \alpha' \qquad (*)$$

Then we have the following result.

Proposition 1: Let $0<\alpha<\alpha'<1$ be such that $(*)$ holds for every estimator sequence $\{T_n\}$. Then for any sequence of confidence sets $\hat{C}$ that is honest over Θ with coverage $1-\alpha$, one has,

$$\liminf_{n\to\infty} \sup_{\theta\in\Theta_1} P_\theta^{(n)}\left(\mathrm{diam}(\hat{C}_n) \geq \varepsilon_n\right) > \alpha'-\alpha$$

Proof: The intuition behind the above result is pretty simple. $(*)$ implies that no estimator is closer to θ more than ε_n. If an honest confidence set had diameter less than ε_n, then one can choose any point within that set as ~~is~~ an estimator of θ that is ~~is~~ at a distance less than ε_n → and thus contradicting $(*)$.

Formally, given a sequence of $(1-\alpha)$-honest confidence $\hat{C}_n$ over Θ, define for each n, $\{T_n\}$ to be an arbitrary point of $\hat{C}_n$. Then for any $\theta \in \Theta_1$,

$$P_\theta^{(n)}(\|T_n - \theta\| \geq \varepsilon_n) \leq P_\theta^{(n)}(\mathrm{diam}(\hat{C}_n) \geq \varepsilon_n) + P_\theta^{(n)}(\theta \notin \hat{C}_n)$$

But $\liminf_{n\to\infty} \inf_{\theta \in \Theta_1} P_\theta^{(n)}(\theta \in \hat{C}_n) \geq 1-\alpha$

& $\liminf_{n\to\infty} \sup_{\theta \in \Theta_1} P_\theta^{(n)}(\|T_n - \theta\| \geq \varepsilon_n) \geq \alpha'$

$\Rightarrow \liminf_{n\to\infty} \sup_{\theta \in \Theta_1} P_\theta^{(n)}(\mathrm{diam}(\hat{C}_n) \geq \varepsilon_n) \geq \alpha' - \alpha$ ∎

(b) Lower Bound from Testing Argument

Let $0 < \alpha < \alpha' < 1$ be given. Let $\{\varepsilon_n\}$ be a sequence of positive real numbers, such that, there are no sequence of tests $\{\varphi_n\}$ satisfying the two requirements below, for some given subset sequence $\Theta_{n,1} \subseteq \Theta_1$,

(i) $\limsup_{n\to\infty} \sup_{\theta \in \Theta_{n,1}} P_\theta^{(n)}(\varphi_n) < \alpha'$

(ii) $\limsup_{n\to\infty} \sup_{\theta \in \Theta : \|\theta, \Theta_{n,1}\| > \varepsilon_n} P_\theta^{(n)}(1-\varphi_n) < \alpha$

Proposition 2: For given $0<\alpha<\alpha'<1$ and subsets $\Theta_{n,1} \subset \Theta_1$, If there exists no sequence of tests $\{\varphi_n\}$ satisfying (i) and (ii) above, then for any sequence of $1-\alpha$ honest confidence set over Θ,

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_1} P_\theta^{(n)}\left(\mathrm{diam}(\hat{C}_n) \geq \varepsilon_n\right) \geq \alpha' - \alpha .$$

Proof: The proof relies on ability of inverting confidence sets to construct statistical tests. Intuitively, if for $\theta\in\Theta_{n,1}$ the $\mathrm{diam}(\hat{C}_n)<\varepsilon_n$ for a honest $\hat{C}_n$, then note that ~~for any~~ ~~are~~ ~~can potentially reject~~ ~~any~~ an ε_n-ball around $\Theta_{n,1}$ should definitely contain ~~intersect~~ $\hat{C}_n$. Therefore one should reject $\theta\in\Theta_{n,1}$ if an ε_n-ball around $\Theta_{n,1}$ does not ~~reject~~ contain ~~intersect~~ $\hat{C}_n$. This test will lead us to a contradiction of (i) and (ii).

Formally, let $\Theta_{n,1}^{\varepsilon} = \{\theta\in\Theta : d(\theta, \Theta_{n,1}) \leq \varepsilon_n\} \rightarrow$ closed set.

Let $\varphi_n = I\{\hat{C}_n \not\subseteq \Theta_{n,1}^{\varepsilon}\}$. Let us study the error properties of the test against

$$H_0: \theta\in\Theta_{n,1} \quad vs \quad H_1: \theta\in\Theta : d(\theta, \Theta_{n,1}) > \varepsilon_n = \left(\Theta_{n,1}^{\varepsilon}\right)^c$$

If $\theta\in\left(\Theta_{n,1}^{\varepsilon}\right)^c$, then

$$P_\theta^{(n)}(\varphi_n = 0) = P_\theta^{(n)}\left(\hat{C}_n \subseteq \Theta_{n,1}^{\varepsilon}\right) \leq P_\theta^{(n)}\left(\theta \notin \hat{C}_n\right)$$

$$\Rightarrow \limsup_{n\to\infty} \sup_{\theta\in\left(\Theta_{n,1}^{\varepsilon}\right)^c} P_\theta^{(n)}(\varphi_n=0) \leq \alpha \quad \text{by honesty.}$$

Therefore one must have $\limsup_{n\to\infty} \sup_{\theta \in \Theta_{n,1}} P_\theta^{(n)}(\varphi_n = 1) \geq \alpha'$

But $P_\theta^{(n)}(\varphi_n = 1) = P_\theta^{(n)}\left(\hat{C}_n \not\subseteq \Theta_{n,1}^{\varepsilon_n}\right)$

$$= P_\theta^{(n)}\left(\theta \in \hat{C}_n,\ \hat{C}_n \not\subseteq \Theta_{n,1}^{\varepsilon_n}\right) + P_\theta^{(n)}\left(\theta \notin \hat{C}_n,\ \hat{C}_n \not\subseteq \Theta_{n,1}^{\varepsilon_n}\right)$$

$$\Rightarrow \sup_{\theta \in \Theta_{n,1}} P_\theta^{(n)}(\varphi_n = 1) \leq \sup_{\theta \in \Theta_{n,1}} P_\theta^{(n)}\left(\mathrm{diam}(\hat{C}_n) \geq \varepsilon_n\right) + \sup_{\theta \in \Theta_{n,1}} P_\theta^{(n)}\left(\theta \notin \hat{C}_n\right)$$

$$\Rightarrow \limsup_{n\to\infty} \sup_{\theta \in \Theta_{n,1}} P_\theta^{(n)}\left(\mathrm{diam}(\hat{C}_n) \geq \varepsilon_n\right) \geq \alpha' - \alpha \quad \blacksquare$$

Construction of confidence sets

We now discuss the construction of confidence sets in the general set up followed by specific example in the density model. The method is based on sample splitting. Suppose that an initial estimator $\hat{\theta}^{(n)}$ of θ is given and we construct confidence sets based on $\hat{\theta}^{(n)}$ and an additional independent observation $X^{(n)}$. The nature of the initial estimator is irrelevant for the sake of honesty & will only come into play while demanding adaptation over submodels.

Our confidence set construction is motivated by Robins & Van Der Vaart (2006) and is based on estimator $R_n(\hat{\theta}^{(n)}, X^{(n)})$ of the squared norm $\|\theta - \hat{\theta}^{(n)}\|^2$ satisfying

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta} P_\theta\left(R_n(\hat{\theta}^{(n)}, X^{(n)}) - \|\theta - \hat{\theta}^{(n)}\|^2 \geq -z_\alpha \hat{\tau}_{n,\theta} \,\Big|\, \hat{\theta}^{(n)}\right) \geq 1-\alpha$$

for "scale-estimators" $\hat{\tau}_{n,\theta}$ and "quantiles" z_α, $0<\alpha<1$. The probability in the above display is computed conditionally on $\hat{\theta}^{(n)}$. By Fatou's Lemma,

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta} P_\theta\left(R_n(\hat{\theta}^{(n)}, X^{(n)}) - \|\theta - \hat{\theta}^{(n)}\| \geq -z_\alpha \hat{\tau}_{n,\theta}\right) \geq 1-\alpha$$

$\Rightarrow$ The set, $\hat{C}_n = \left\{\theta\in\Theta : \|\theta - \hat{\theta}^{(n)}\| \leq \sqrt{z_\alpha \hat{\tau}_{n,\theta} + R_n(\hat{\theta}^{(n)})}\right\}$ is a $1-\alpha$ honest confident set over Θ. This confidence set is not necessarily a ball. However, in most examples,

$$\hat{\tau}_{n,\theta} \leq C(\Theta)\left\{\hat{\tau}_n + \frac{\|\theta - \hat{\theta}^{(n)}\|}{\sqrt{n}}\right\}$$

where $\hat{\tau}_n$ is decided by size of Θ. Then we can take

$$\hat{C}_n = \left\{\theta\in\Theta : \|\theta - \hat{\theta}^{(n)}\| \leq \sqrt{C(\Theta)\left\{\hat{\tau}_n + \frac{\|\theta - \hat{\theta}^{(n)}\|}{\sqrt{n}}\right\} z_\alpha + R_n(\hat{\theta}^{(n)})}\right\}$$

which is now a ball with diameter,

$$\operatorname{diam}(\hat{C}_n) \lesssim \sqrt{\hat{\tau}_n} + \sqrt{R_n(\hat{\theta}^{(n)})} + n^{-1/2}$$

We shall see that the following typically happens:

(i) $n^{-1/2}$ being parametric rate is typically negligible.

(ii) $\sqrt{\hat{\tau}_n}$ typically depends on Θ with size being similar over submodels.

(iii) The possibility of adaptation relies on $\sqrt{R_n(\hat{\theta}^{(n)}, X_n)}$.

The Crucial Term: $R_n(\hat\theta^{(n)}, X^{(n)})$ typically will satisfy

$$\left|R_n(\hat\theta^{(n)}, X^{(n)}) - \|\theta - \hat\theta^{(n)}\|^2\right| = O_{P_\theta}(\hat\tau_{n,\theta})$$

uniformly in $\theta \in \Theta$. $\leftarrow$ standard deviation.

$\Rightarrow \operatorname{diam}(\hat C_n) = O_{P_\theta}\left(\sqrt{\hat\tau_n} + \|\theta - \hat\theta^{(n)}\| + n^{-1/2}\right)$ uniformly in $\theta \in \Theta$.

$\Rightarrow \operatorname{diam}(\hat C_n)$ for $\theta \in \Theta_1 \subset \Theta$ is bounded above by largest term (in order) of the RHS above under θ for $\theta \in \Theta_1$. Typically for small submodels, $\sqrt{\hat\tau_n}$ will dominate (if $\hat\theta^{(n)}$ performs well, $\hat\tau_n$ being the rate of estimation of $\|\theta - \hat\theta^{(n)}\|_2^2$. For bigger submodels on the other hand, $\|\theta - \hat\theta^{(n)}\|$ dominates $\rightarrow$ the rate of estimation of θ. (This reminds us between the testing rate/estimation rate balance/trade off)

Note: Estimation of $\|\theta - \hat\theta^{(n)}\|^2$ is basically estimation of a quadratic functional. We can do this by our standard projection onto finite dimensional subspace type argument as follows.

We estimate $\|\Pi_k\theta - \Pi_k\hat\theta^{(n)}\|_2^2$ instead where Π_k denotes the projection onto $\{e_1, \dots, e_k\}$ with $\{e_i\}_{i \geq 1}$ being an o.n.b. of the Hilbert space. We can estimate this unbiasedly & then trade off the resulting squared bias for estimating $\|\theta - \hat\theta^{(n)}\|^2$ against the variance of the estimator.

If $\hat{\theta}^{(n)} \in \Theta$, then the (bias)[*] is (in absolute value)

$$B_k^2 \leq \sup_{\theta' \in \Theta} \|\theta' - \Pi_k \theta'\|^2$$

The variance turns out to be of the order, for a parameter σ^2 (that depends on the problem)

$$\hat{\tau}^2_{k,n,\theta} = \frac{2\sigma^4 k}{n^2} + \frac{4\sigma^2 \|\Pi_k(\theta - \hat{\theta}^{(n)})\|^2}{n}$$

Therefore $\hat{\tau}_{n,\theta} = \dfrac{\sqrt{2}\,\sigma^2\sqrt{k}}{n} + 2\sigma \dfrac{\|\Pi_k(\theta - \hat{\theta}^{(n)})\|}{\sqrt{n}}$

$$+ B_k^2$$

$$\leq \frac{\sqrt{2}\,\sigma^2\sqrt{k}}{n} + \frac{2\sigma\,\|\theta - \hat{\theta}^{(n)}\|}{\sqrt{n}} + B_k^2\left(1 + \frac{2\sigma}{\sqrt{n}}\right)$$

$$= \left(\hat{\tau}_n + \frac{2\sigma\,\|\theta - \hat{\theta}^{(n)}\|}{\sqrt{n}}\right)$$

$$\hat{\tau}_n = \frac{\sqrt{2}\,\sigma^2\sqrt{k}}{n} + B_k^2\left(1 + 2\sigma/\sqrt{n}\right)$$

$\Rightarrow$ Resulting diameter,

$$\operatorname{diam}(\hat{C}_n) \lesssim \frac{\sigma k^{1/4}}{\sqrt{n}} + B_k + \|\theta - \hat{\theta}^{(n)}\| + \sigma/\sqrt{n}$$

We now choose k by trading off $k^{1/4}/\sqrt{n}$ with B_k. (The parameter σ may depend on θ, and must be bound uniformly over Θ in that case)

Let us now formalize this discussion.

Proposition 3: Let us assume the following for $k \in \mathbb{N}$, $n \in \mathbb{N}$

(i) $\hat{\tau}_{k,n,\theta} = \sqrt{2}\,\sigma^2 \sqrt{k}/n + B_k^2$ for $\sigma \in (0, \bar{\sigma})$, $B_k = \sup_{\theta \in \Theta} \|\theta - \Pi_k \theta\|^2$

For some $z_\alpha > 0$, $k_n \to \infty$ and $M_n \to \infty$ being any sequences.

(ii) $\limsup_{n \to \infty} \sup_{\theta \in \Theta} P_\theta^{(n)}\left(R_{k_n,n}(\hat{\theta}^{(n)}, X(n)) - \|\Pi_{k_n}\theta - \Pi_{k_n}\hat{\theta}^{(n)}\|^2 \leq -z_\alpha \hat{\tau}_{k_n,n,\theta}\right) \leq \alpha$

(iii) $\limsup_{n \to \infty} \sup_{\theta \in \Theta} P_\theta^{(n)}\left(\left|R_{k_n,n}(\hat{\theta}^{(n)}, X(n)) - \|\Pi_{k_n}\theta - \Pi_{k_n}\hat{\theta}^{(n)}\|^2\right| \geq M_n \hat{\tau}_{k_n,n,\theta}\right) = 0$

where $R_{k,n}(\hat{\theta}^{(n)}, X(n))$ are above discussed estimator of $\|\Pi_{k_n}\theta - \Pi_{k_n}\hat{\theta}^{(n)}\|^2$.

If θ takes values in Θ and

$$\hat{C}_n = \left\{\theta : \|\theta - \hat{\theta}^{(n)}\| \leq \sqrt{z_\alpha \hat{\tau}_{k_n,n,\theta} + R_{k_n,n}(\hat{\theta}^{(n)}, X(n))} + 2B_{k_n}\right\},$$

then $\hat{C}_n$ is (1-α) honest over $\theta \in \Theta$, with diameter satisfying,

$$\limsup_{n \to \infty} \sup_{\theta \in \Theta} P_\theta^{(n)}\left(\mathrm{diam}(\hat{C}_n) \geq M_n\left[\frac{\bar{\sigma}\, k^{1/4}}{\sqrt{n}} + B_{k_n} + \|\theta - \hat{\theta}^{(n)}\|\right]\right) = 0$$

Proof: Since $\|\theta - \hat{\theta}^{(n)}\| \leq \|\Pi_{k_n}\theta - \Pi_{k_n}\hat{\theta}^{(n)}\| + 2B_{k_n}$

$$\Rightarrow P_\theta^{(n)}(\theta \notin \hat{C}_n) \leq P_\theta^{(n)}\left(\|\theta - \hat{\theta}^{(n)}\| - 2B_{k_n} > \sqrt{z_\alpha \hat{\tau}_{k_n,n,\theta} + R_{k_n,n}(\hat{\theta}^{(n)}, X(n))}\right)$$

$$\leq P_\theta^{(n)}\left(\|\Pi_k\theta - \Pi_k\hat{\theta}^{(n)}\|^2 > z_\alpha \hat{\tau}_{k_n,n,\theta} + R_{k_n,n}(\hat{\theta}^{(n)}, X(n))\right)$$

$$= P_\theta^{(n)}\left(R_{k_n,n}(\hat{\theta}^{(n)}, X(n)) - \|\Pi_k\theta - \Pi_k\hat{\theta}^{(n)}\|^2 < -z_\alpha \hat{\tau}_{k_n,n,\theta}\right)$$

$\Rightarrow$ By (ii) $\limsup_{n \to \infty} \sup_{\theta \in \Theta} P_\theta^{(n)}(\theta \notin \hat{C}_n) \leq \alpha$. $\to$ Honesty!

Now by definition of $\hat{\tau}_{k_n, n, \theta}$, every $\theta \in \hat{C}_n$ satisfies

$$\|\theta - \hat{\theta}^{(n)}\| \leq \sqrt{z_\alpha \frac{\sqrt{2}\sigma^2}{n}\sqrt{k_n} + R_{k_n,n}(\hat{\theta}^{(n)}, X^{(n)}) + 2B_{k_n}} + \frac{\sqrt{z_\alpha 2\sigma}}{n^{1/4}} \sqrt{\|\theta - \hat{\theta}^{(n)}\|}.$$

($\sqrt{a} + \sqrt{b} \geq \sqrt{a+b}$ for $a, b \geq 0$). Therefore the diameter of $\hat{C}_n$ is bounded by a multiple of (using $\sigma \in (0, \bar{\sigma})$)

$$\frac{\bar{\sigma} k^{1/4}}{\sqrt{n}} + \sqrt{R_{k_n,n}(\hat{\theta}^{(n)}, X^{(n)})} + B_{k_n} + \frac{\bar{\sigma}}{\sqrt{n}}$$

(since, $x \leq a + b\sqrt{x} \Rightarrow x \leq 2a + 2b^2$ if $a, b \geq 0$)

The variable $R_{k_n,n}(\hat{\theta}^{(n)}, X^{(n)})$ are with $P_\theta^{(n)}$- probability tending to 1 bounded above by a multiple of $\|\Pi_{k_n}\theta - \Pi_{k_n}\hat{\theta}^{(n)}\|^2 + M_n \hat{\tau}_{k_n,n,\theta}$ for any given sequence $M_n \to \infty$

$\Rightarrow$ With probability tending to 1, the diameter of $\hat{C}_n$ is bounded by a multiple of

$$\frac{\bar{\sigma} k^{1/4}}{\sqrt{n}} + \|\theta - \hat{\theta}^{(n)}\| + B_{k_n} + \sqrt{M_n \hat{\tau}_{k_n,n,\theta}} + \frac{\bar{\sigma}}{\sqrt{n}}$$

The last term above is negligible to the first.

Now,

$$\hat{\tau}^2_{k_n,n,\theta} = \frac{2\sigma^4}{n^2} k + \frac{4\sigma^2}{n} \|\Pi_{k_n}\theta - \Pi_{k_n}\hat{\theta}^{(n)}\|^2$$

$$\leq \frac{2\sigma^4}{n^2} k + \frac{4\sigma^2}{n} \|\theta - \hat{\theta}^{(n)}\|^2$$

$$\leq \frac{2\bar{\sigma}^4}{n^2} k + \frac{4\bar{\sigma}^2}{n} \|\theta - \hat{\theta}^{(n)}\|^2$$

$\therefore$ Diameter of $\hat{C}_n$ is bounded by a multiple of

$$\frac{\bar{\sigma}\, k_n^{1/4}}{\sqrt{n}} + \|\theta - \hat{\theta}^{(n)}\| + \left(\frac{\bar{\sigma}\, k_n^{1/4}}{\sqrt{n}} + \frac{\sqrt{\bar{\sigma}}\sqrt{\|\theta - \hat{\theta}^{(n)}\|}}{n^{1/4}}\right)\sqrt{M_n}\; \cancel{\frac{\sqrt{\bar{\sigma}}\sqrt{B_k}}{n^{1/4}}}$$

$$+ B_{k_n}$$

$$= \left(B_{k_n} + 2\bar{\sigma}\frac{k_n^{1/4}}{\sqrt{n}} + \frac{\sqrt{\bar{\sigma}}}{n^{1/4}}\sqrt{\|\theta - \hat{\theta}^{(n)}\|} + \|\theta - \hat{\theta}^{(n)}\|\right)\sqrt{M_n}$$

Now, $\dfrac{\bar{\sigma}\, k^{1/4}}{\sqrt{n}} + \dfrac{\sqrt{\bar{\sigma}}}{n^{1/4}}\sqrt{x} + x \leq \dfrac{2\bar{\sigma}\, k^{1/4}}{\sqrt{n}} + 2x$ for any $x \geq 0$, $k \geq 1$ and $\bar{\sigma} > 0$

$\Rightarrow$ Diameter of $\hat{C}_n$ is bounded by a multiple of (with $P_\theta^{(n)}$- probability $\to 1$ uniformly under $\theta \in \Theta$)

$$M_n\left(\frac{\bar{\sigma}\, k_n^{1/4}}{\sqrt{n}} + B_{k_n} + \|\theta - \hat{\theta}^{(n)}\|\right)$$ as required ∎