Lecture 14

In this lecture we will try to understand construction of adaptive confidence sets for our usual density model with diameter measured in the L_2-norm. We will see that a natural space of functions to look at for the purpose of this adaptive confidence building is $B^{\beta}_{2\infty}(M)$ defined in terms of wavelet coefficients as follows.

Fix a CDV family of wavelet basis on $L_2[0,1]$ s.t. (φ, ψ) are C^r with $r \geq \lceil \beta \rceil$. Then define

$$B^{\beta}_{2\infty}(M) := \left\{ f \in L_2[0,1] : \sup_{l \geq J_0} 2^{l\beta} \left(\| \langle f, \psi_{l\cdot} \rangle \|_2 \right) \leq M \right\}$$

where J_0 is typically made to depend on (φ, ψ) chosen (for technical reasons that we won't specify here).

In the context of this function space, we will be interested in the following set up as usual.

Set up: $X_1, \ldots, X_n \overset{iid}{\sim} P$ on $([0,1], \mathcal{B}[0,1])$ with $\frac{dP}{d\lambda} = \eta$ (λ being the Lebesgue measure on $[0,1]$). Let $0 < \beta_{min} < \beta_{max} < \infty$ be given along with $M, B > 0$. We want to produce $(1-\alpha)$ honest adaptive confidence sets over $\bigcup_{\beta \in [\beta_{min}, \beta_{max}]} \mathcal{P}(\beta, M, B)$ where

$$\mathcal{P}(\beta, M, B) := \left\{ f \in L_2[0,1] : 0 \leq f \leq B \text{ a.e. } \lambda, \int f \, d\lambda = 1, f \in B^{\beta}_{2\infty}(M) \right\}$$

with $B^{\beta}_{2\infty}(M)$ defined using (φ, ψ) CDV type and regularity $r \geq \lceil \beta_{max} \rceil$.

It turns out that this problem also has the same two regions of interest.

(i) Adaptation Region $(\beta_{max} \leq 2\beta_{min})$

(ii) Non Adaptation Region $(\beta_{max} > 2\beta_{min})$

(i) The Adaptation Region. In this region one can prove an analogue of the results we obtained in lecture 12. We collect these in the next theorem.

Assume $\beta_{max} \leq 2\beta_{min}$

Theorem 1: There exists a measurable set $\hat{C}_n \subseteq (L_2[0,1], \mathcal{B}(L_2[0,1]))$ based on $(X_1, \ldots, X_n, M, B, \alpha, \beta_{min}, \beta_{max})$ such that the following hold.

(a) $$\liminf_{n\to\infty} \inf_{\eta \in \bigcup_{\beta\in[\beta_{min},\beta_{max}]} \mathcal{P}(\beta,M,B)} \mathbb{P}_\eta(\eta \in \hat{C}_n) \geq 1-\alpha.$$

(b) For any $\beta \in [\beta_{min}, \beta_{max}]$,

$$\limsup_{n\to\infty} \sup_{\eta\in\mathcal{P}(\beta,M,B)} \frac{\mathbb{E}_\eta(\mathrm{diam}(\hat{C}_n))}{n^{-\beta/2\beta+1}} < \infty.$$

Remark: This is called the Adaptation Region because the rate of estimating the whole density η over any $\mathcal{P}(\beta,M,B)$ ~~is~~ in the L_2-norm is $n^{-\beta/2\beta+1}$. The proof of Theorem 1 uses a similar technique of risk estimation. For more details see: Adaptive Confidence Sets in L_2: Richard Nickl & Adam Bull (Probability Theory & Related Fields 2013)

(ii) Ihe Non-Adaptation Region: We will however be more focused on the non-adaptive region i.e. $\beta_{max} > 2\beta_{min}$. Here we shall show using our testing lower bound argument that one cannot have honest confidence sets $\{\hat{C}_n\}$ over $\bigcup_{\beta \in [\beta_{min}, \beta_{max}]} \mathcal{P}(\beta, M, B)$ which shrink at a rate of $n^{-\beta/2\beta+1}$ uniformly over all $\mathcal{P}(\beta, M, B)$ for each $\beta \in [\beta_{min}, \beta_{max}]$ if $\beta_{max} > 2\beta_{min}$. (This is too large a collection of parameter space to honestly adapt over). To this end we state the next theorem, which is interesting in its own right.

Theorem 2: Let $\beta_{max} > \beta_{min}$ be given. Consider the hypothesis testing problem

$$H_0: \eta \in \mathcal{P}(\beta_{max}, M, B) \quad \text{vs} \quad H_1^{(\rho_n)}: \{\eta \in \mathcal{P}(\beta_{min}, M, B) : \|\eta - \eta'\| \geq \rho_n \text{ for all } \eta' \in \mathcal{P}(\beta_{max}, M, B)\}$$

Then

(a) If $\rho_n^2 \gg n^{-\frac{4\beta_{min}}{4\beta_{min}+1}}$, $\exists$ a test $\{\varphi_n\}$ such that

$$\limsup_{n\to\infty} \left[\sup_{\eta \in H_0} \mathbb{P}_\eta(\varphi_n = 1) + \sup_{\eta \in H_1(\rho_n)} \mathbb{P}_\eta(\varphi_n = 0) \right] = 0$$

(b) If $\rho_n^2 \ll n^{-\frac{4\beta_{min}}{4\beta_{min}+1}}$, for every test sequence $\{\varphi_n\}$

$$\liminf_{n\to\infty} \left[\sup_{\eta \in H_0} \mathbb{P}_\eta(\varphi_n = 1) + \sup_{\eta \in H_1(\rho_n)} \mathbb{P}_\eta(\varphi_n = 0) \right] = 1.$$

Proof: We prove (b) first since the idea of the proof is similar to what we saw before. The proof of (a) follows an idea due to Carpentier (2015).

(b) The proof relies on noting that the rate of testing does not seem to depend on the complexity of the null hypothesis. This in turn implies that each element in the null hypothesis has an equal footing of difficulty in the problem. So we can simply look at a less difficult problem with a simple null hypothesis, prove the same lower bound rate in the simpler problem, and conclude that the original problem has atleast the same level of difficulty. Mathematically for any test sequence $\{\varphi_n\}$,

$$\sup_{\eta \in H_0} \mathbb{P}_\eta(\varphi_n = 1) + \sup_{\eta \in H_1(\rho_n)} \mathbb{P}_\eta(\varphi_n = 0)$$

$$\geq \mathbb{P}_{\eta_0}(\varphi_n = 1) + \sup_{\eta \in H_1(\rho_n)} \mathbb{P}_\eta(\varphi_n = 0)$$

for any $\eta_0 \in H_0$. Take $\eta_0 \equiv 1$ from now on.

$$\geq (1-\varepsilon)\left(1 - \sqrt{\chi^2\left(P^n_{\eta_0}, (P^n)_\pi\right)/\varepsilon}\right)$$

for any $\varepsilon > 0$ where

$$\chi^2(P, Q) = \mathbb{E}_P\left(\left(\frac{dQ}{dP}\right)^2\right) - 1$$

$$P^n_\pi(A) = \int_{H_1(\rho_n)} \int_A dP^n_\eta \, d\pi$$

(slight abuse of notation since here by $H_1(\rho_n)$ we mean $\{P^n_\eta : \eta \in H_1(\rho_n)\}$)

The last inequality holds for any $\varepsilon > 0$ and any prior π supported on the alternative. We construct one such prior very similar to before.

Take CDV type wavelets of sufficient regularity ($r \geq \lceil \beta_{max} \rceil$). Recall that at resolution level j, take the wavelets ψ_{jk} whose supports are inside $[0,1]$. ~~For~~ This number of wavelets is bounded by $C(\psi) 2^j$. We denote by $\mathcal{Z}_j$ the corresponding index sets k and hence $|\mathcal{Z}_j| \lesssim C(\psi) 2^j$. Indeed we can take for j large enough, $|\mathcal{Z}_j| = C(\psi) 2^j$ for some $C(\psi)$. For $\underset{\sim}{\lambda} = (\lambda_k, k \in \mathcal{Z}_j) \in \{-1,+1\}^{|\mathcal{Z}_j|}$ and $\varepsilon > 0$ a small constant, let

$$\eta_{\underset{\sim}{\lambda}}(x) = \eta_0(x) + \varepsilon 2^{-j\beta_{min}} \cdot 2^{-j/2} \sum_{k \in \mathcal{Z}_j} \lambda_k \psi_{jk}(x) \qquad \underset{\sim}{\lambda} \in \{-1,+1\}^{|\mathcal{Z}_j|}$$

We know by regularity $\Rightarrow$ vanishing moments that all the ψ_{jk}'s integrate to 0 on $[0,1]$. Also,

$$\|\eta_\lambda - \eta_0\|_\infty \leq \varepsilon 2^{-j\beta_{min}} \sup_{x \in \mathbb{R}} \sum_{k \in \mathbb{Z}} |\psi(2^j x - k)| \leq C(\psi, \varepsilon) 2^{-j\beta_{min}}$$

$\Rightarrow$ For ε small enough $\eta_{\underset{\sim}{\lambda}}$ ~~$\in$~~ are uniformly bounded probability densities. Especially for $B > 1$, they are still bounded by B. And we can assume $B > 1$ w.l.o.g.

Now, the Besov norm of η_0 is 1. For the η_λ's we have by the wavelet characterization of the Besov spaces and the fact that the ψ_{jk}'s chosen have support inside $[0,1]$ that

$$\|\eta_{\underset{\sim}{\lambda}} - \eta_0\|^2_{B^{\beta_{min}}_{2\infty}[0,1]} = \varepsilon^2 2^{2j\beta_{min}} 2^{-2j\beta_{min}} 2^{-j} \sum_{k \in \mathcal{Z}_j} \lambda_k^2 \lesssim C(\psi) \varepsilon^2$$

$\Rightarrow$ For ε small enough, $\underset{\sim}{\eta_\lambda} \in \mathcal{P}(\beta_{min}, M, B)$.

Also $\|\eta_\lambda - \eta_0\|_2^2 = \varepsilon^2 2^{-2j\beta_{min}}$ Now choose ~~$2^{j\beta}$~~
$2^{-2j\beta_{min}} = o(\rho_n^2)$ which
mean $2^j \gg n^{\frac{2}{4\beta_{min}+1}}$

The rest of the proof for controlling the χ^2-divergence between the null and the mixture of alternatives is same as the proof of the lower bound theorem in Lecture 6. $\blacksquare$

(a) As mentioned before the proof of the upper bound is based on an idea from the paper: "Testing Regularity of a Smooth Signal: Alexandra Carpentier, Bernoulli, 2015.
To explain the idea, let

For $j \geq 0$, $W_j = \text{span}\{\psi_{jk},\ k \in \mathcal{Z}_j\}$ $\quad \mathcal{Z}_j \equiv \{k$ s.t. support of ψ_{jk} intersects $[0,1]\}$.

$\Pi_{W_j}(f) \equiv$ projection of $f \in L_2[0,1]$ onto the space W_j, $j \geq 0$.

(use CDV type basis which guarantees $|\mathcal{Z}_j| = c(\psi)\, 2^j$).

$$\Pi_{W_j}(f) = \int K_{W_j}(\cdot, y)\, f(y)\, dy$$

$$K_{W_j}(x,y) = \sum_{k \in \mathcal{Z}_j} \psi_{jk}(x)\, \psi_{jk}(y).$$

The idea is ~~the~~ based on an observation that under the null hypothesis $\|\Pi_{W_j}(\eta)\|_2^2$ is not too large for $J_0 \leq j \leq j^*$ (J₀ comes from the construction of CDV wavelets i.e $f(x) = \sum_{k \in \mathcal{Z}_{J_0}} \alpha_{J_0 k}\, \varphi_{J_0 k}(x) + \sum_{l \geq J_0} \sum_{k \in \mathcal{Z}_{J_0}} \beta_{lk}\, \psi_{lk}(x)$)
for j^* to be decided later. Whereas under the alternative $\exists$ a resolution j for which $\|\Pi_{W_j}(\eta)\|_2^2$ is large.

Therefore one implements a scan type argument, where we proceed by estimating $\|\Pi_{W_j}(\eta)\|_2^2$ for each j and reject if any one of them is "large". We now do this more formally.

<u>Estimation of $\|\Pi_{W_\ell}(\eta)\|_2^2$</u>

This is easy to do using $\frac{1}{n(n-1)}\sum_{i\neq j} K_{W_\ell}(X_i, X_j)$. The crucial part is the analysis of this estimator.

Let $T_n(\ell) = \frac{1}{n(n-1)}\sum_{i\neq j} K_{W_\ell}(X_i, X_j)$

Its easy to check that $\mathbb{E}_\eta(T_n(\ell)) = \|\Pi_{W_\ell}(\eta)\|_2^2$.

$$\mathrm{Var}_\eta(T_n(\ell)) \leq C(\psi, M, B)\left\{\frac{\|\Pi_{W_\ell}(f)\|_2^2}{n} + \frac{2^\ell}{n^2}\right\}$$ (check this by usual Hoeffding's Decomposition)

Therefore $T_n(\ell)$'s are unbiased estimators of $\|\Pi_{W_\ell}(\eta)\|_2^2$ and have fluctuations of the order of $\left\{\frac{\|\Pi_{W_\ell}(f)\|_2^2}{n} + \frac{2^\ell}{n^2}\right\}^{\frac{1}{2}}$, and we know $\|\Pi_{W_\ell}(\eta)\|_2^2$ is "small" under the null.

If we were just testing at a single level we could have rejected if

$$T_n(\ell) \geq \text{"}\|\Pi_{W_\ell}(\eta)\|_2^2\text{"} + C^*\left\{\frac{2^\ell}{n^2} + \frac{\|\Pi_{W_\ell}(f)\|_2^2}{n}\right\}$$

where "$\|\Pi_{W_\ell}(\eta)\|_2^2$" is used to mean some quantity depending on the typical size of $\|\Pi_{W_\ell}(\eta)\|_2^2$ under the null hypothesis.

But we want to a list at each level $J_0 \le l \le j^*$ where j^* will be used to diverge with sample size n. so we need to blowup the cut-off / fluctuations to adjust for multiple listing. This is captured by the next lemma.

<u>lemma</u>: $\exists\ C(\underset{\wedge}{\varepsilon,}\psi, M, B) > 0$ sufficiently large such that for $\varepsilon > 0$

$$\mathbb{P}_\eta\left(|T_n(l) - \|\Pi_{W_l}(\eta)\|_2^2 \le C(\varepsilon, \psi, M, B)\sqrt{\frac{2^{(l+J_0)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l}(\eta)\|_2^2}{n}} \quad \forall\ J_0 \le j \le j^* \right) \ge 1 - \frac{\varepsilon}{2}$$

<u>Proof</u>: The proof is simple and is based on union bound and Chebychev's Inequality.

First note that, for any l, by Chebychev's Inequality,

$$\mathbb{P}_\eta\left(|T_n(l) - \|\Pi_{W_l}(\eta)\|_2^2| > C\sqrt{\frac{2^{(l+j_*)/2}}{n^2} + 2^{l/4}\frac{\|\Pi_{W_l}(\eta)\|_2^2}{n}}\right)$$

$$\le \frac{C(\psi, M, B)\left\{\frac{2^l}{n^2} + \|\Pi_{W_l}(\eta)\|_2^2/n\right\}}{C^2\left\{\frac{2^{(l+j_*)/2}}{n^2} + \frac{2^{l/4}\|\Pi_{W_l}(\eta)\|_2^2}{n}\right\}}$$

$$\le \frac{C(\psi, M, B)}{C^2}\left[\frac{\left\{\frac{2^l}{n^2} \cancel{+ \|\Pi_{W_l}(\eta)\|_2^2/n}\right\}}{\left\{\frac{2^{(l+j_*)/2}}{n^2}\right\}} + \frac{\left\{\|\Pi_{W_l}(\eta)\|_2^2/n\right\}}{\left\{\frac{2^{l/4}\|\Pi_{W_l}(\eta)\|_2^2}{n}\right\}}\right]$$

$$\le \frac{C(\psi, M, B)}{C^2}\left[2^{-(j_*-l)/2} + 2^{-l/4}\right]$$

$\therefore$ For any $c > 0$, by union bound,

$$\mathbb{P}_\eta\left(\exists\, J_0 \le \ell \le j_* : \left| T_n(\ell) - \|\Pi_{W_\ell}(\eta)\|_2^2 \right| > c\sqrt{\frac{2^{(\ell+j_*)/2}}{n^2} + 2^{\ell/4}\frac{\|\Pi_{W_\ell}(\eta)\|_2^2}{n}}\right)$$

$$\le \frac{C(\Psi, M, B)}{c^2}\left(\sum_{\ell=J_0}^{j_*} 2^{-j_*/2}\cdot 2^{\ell/2} + \sum_{\ell=J_0}^{j_*} 2^{-\ell/4}\right)$$

$$\le \frac{C(\Psi, M, B)}{c^2}\left(2^{-j^*/2}\frac{2^{j^*/2}-1}{(\sqrt{2}-1)} + \frac{1-2^{-j^*/4}}{1-2^{-1/4}}\right)$$

$$\le \frac{C(\Psi, M, B)}{c} \times \frac{1}{(2^{\frac{1}{4}}-1)} \times \left[1+2^{1/4}\right] \le \varepsilon$$ choosing c sufficiently large depending on ε, Ψ, M, B.

This completes the proof of the lemma. ■

Now we need to use this lemma to construct a multiple testing/scan type of procedure. As mentioned earlier, this depends on the size of $\|\Pi_{W_\ell}(\eta)\|_2^2$ when $\eta \in H_0 = \mathcal{P}(\beta_{max}, M, B)$. To this end note that

$V_\ell = V_{J_0} \oplus W_{J_0} \oplus W_{J_0+1} \oplus \cdots \oplus W_\ell = V_{\ell-1} \oplus W_\ell$

$\therefore \|\Pi_{V_\ell}(\eta)\|_2^2 = \|\Pi_{V_{\ell-1}}(\eta)\|_2^2 + \|\Pi_{W_\ell}(\eta)\|_2^2$

$\Rightarrow \|\Pi_{V_{\ell-1}^\perp}(\eta)\|_2^2 = \|\Pi_{W_\ell}(\eta)\|_2^2 + \|\Pi_{W_{\ell+1}}(\eta)\|_2^2 + \cdots$

But $\eta \in \mathcal{P}(\beta_{max}, M, B) \Rightarrow \|\Pi_{V_{\ell-1}^\perp}(\eta)\|_2^2 \le M/2^{\ell\beta_{max}}$.

$\Rightarrow \|\Pi_{W_\ell}(\eta)\|_2^2 \le \left(M/2^{\ell\beta_{max}}\right)^2$.

Now we know, for all $J_0 \le l \le j_*$ w.p. $\geq 1-\varepsilon/2$,

$$T_n(l) \le \|\Pi_{W_l}(\eta)\|_2^2 + C(\varepsilon,\Psi,M,B)\left\{\frac{2^{(l+j_*)/2}}{n^2} + \frac{2^{l/4}\|\Pi_{W_l}(\eta)\|_2^2}{n}\right\}^{\frac{1}{2}}$$

$$\le \|\Pi_{W_l}(\eta)\|_2^2 + C(\varepsilon,\Psi,M,B)\frac{2^{(l+j_*)/4}}{n} + C(\varepsilon,\Psi,M,B)\, 2^{l/8}\frac{\|\Pi_{W_l}(\eta)\|_2}{\sqrt{n}}$$

$$\le \cancel{\left(\frac{M}{2^{l\beta_{max}}}\right)^{2}} \|\Pi_{W_l}(\eta)\|_2^2 + C^2(\varepsilon,\Psi,M,B)\left(\frac{2^{(l+j_*)/8}}{\sqrt{n}}\right)^2$$

$$+ 2\,C(\varepsilon,\Psi,M,B)\frac{2^{(l+j_*)/8}}{\sqrt{n}}\|\Pi_{W_l}(\eta)\|_2$$

$$\le \left(\|\Pi_{W_l}(\eta)\|_2 + C(\varepsilon,\Psi,M,B)\frac{2^{(l+j_*)/8}}{\sqrt{n}}\right)^2 \qquad \text{if } C(\varepsilon,\Psi,M,B) \geq 1$$

$$\le \left(\frac{M}{2^{l\beta_{max}}} + C(\varepsilon,\Psi,M,B)\frac{2^{(l+j_*)/8}}{\sqrt{n}}\right)^2 := (t_n(l))^2$$

Therefore we reject the null hypothesis at level $\varepsilon/2$ if $\exists\, l \in [J_0, j_*]$ such that $T_n(l) > (t_n(l))^2$.

The crucial question is now analysis of the type II error of this test when $\|\eta-\eta'\|_2^2 \geq \rho_n^2 = n^{-4\beta_{min}/4\beta_{min}+1}$.

Analysis of the Type II error

As we mentioned in the beginning, the crux of the argument relies on the fact that under the alternative $\exists$ an $J_0 \le l \le j_*$ such that $\|\Pi_{W_l}(\eta)\|_2^2$ is large enough and the corresponding $T_n(l)$ beats the cut-off $(t_n(l))^2$ of the test constructed above. This is captured in the next lemma.

Lemma: Let $(\tau_\ell)_{J_0 \le \ell \le j_*}$ be any sequence of positive real numbers. Assume that

$$\rho_n \geq \left(4 \frac{M}{\sqrt{1-2^{-2\beta_{min}}}} 2^{-j_* \beta_{min}} + \frac{4}{3} \sum_{J_0 \le \ell \le j_*} \tau_\ell \right)$$

Then we have, for any $\eta \in H_1(\rho_n)$,

$$\max_{J_0 \le \ell \le j_*} \left(\| \Pi_{W_\ell}(\eta) \|_2 - \frac{M}{2^{\ell \beta_{max}}} - \tau_\ell \right) > 0 \quad \text{if } 2^{j_*} \sim n^{\frac{2}{4\beta_{min}+1}}$$

Proof: Assume $\eta \in H_1(\rho_n)$. Then by triangle inequality,

$$\rho_n \le \inf_{\eta' \in H_0} \| \eta - \eta' \|_2 \le \inf_{\eta' \in H_0} \| \Pi_{V_{j_*}}(\eta) - \eta' \|_2 + \| \eta - \Pi_{V_{j_*}}(\eta) \|_2$$

$$\le \inf_{\eta' \in H_0} \| \Pi_{V_{j_*}}(\eta) - \eta' \|_2 + \frac{M}{\sqrt{1-2^{-2\beta_{min}}}} 2^{-j_* \beta_{min}}$$

since by definition of our Besov Ball, $f \in \mathcal{B}(\beta_{min}, M, B)$

$$\Rightarrow \| f - \Pi_{V_{j^*}}(f) \|_2 \le \sqrt{\sum_{\ell = j^*+1}^{\infty} 2^{-2\ell \beta_{min}} M^2} \le \frac{M}{\sqrt{1-2^{-2\beta_{min}}}} 2^{-j^* \beta_{min}}$$

$$\Rightarrow \inf_{\eta' \in H_0} \| \Pi_{V_{j^*}}(\eta) - \eta' \|_2 \geq \rho_n - \frac{M\, 2^{-j_* \beta_{min}}}{\sqrt{1-2^{-2\beta_{min}}}} \geq 3\rho_n / 4$$

Let us write $(a_{\ell k})_{\ell, k}$ to be the wavelet coefficients of η and $(b_{\ell k})_{\ell, k}$ the wavelet coefficients of the minimizer η'. Then the following string of inequalities can be justified by triangle inequality & Parseval's identity.

$$\inf_{\eta' \in H_0} \| \Pi_{V_{j*}}(\eta) - \eta' \|_2 \leq \inf_{\eta' \in H_0} \sum_{l=J_0}^{j*} \| \Pi_{W_l}(\eta) - \eta' \|_2 \qquad \left(s = \beta_{max} \text{ for shorter notation} \right)$$

$$= \inf_{(b_{lk}) : \forall l \geq J_0,\, 2^{ls} \| b_{l,\cdot} \|_2 \leq M} \left(\sum_{l=J_0}^{j*} \sqrt{\sum_{k \in Z_l} (a_{lk} - b_{lk})^2} + \sum_{l=j*+1}^{\infty} \sqrt{\sum_{k \in Z_l} b_{lk}^2} \right)$$

$$= \sum_{l=J_0}^{j*} \inf_{(b_{lk}) : 2^{ls} \| b_{l\cdot} \|_2 \leq M} \left(\sqrt{\sum_{k \in Z_l} (a_{lk} - b_{lk})^2} \right)$$

$$\leq \sum_{l=J_0}^{j*} \left(0, \sqrt{\sum_{k \in Z_l} a_{lk}^2} - \frac{M}{2^{ls}} \right) \leq \| \Pi_{W_l}(\eta) \|_2 - \frac{M}{2^{l\beta_{max}}}$$

$$\Rightarrow \frac{3\rho_n}{4} \leq \sum_{l=J_0}^{j*} \| \Pi_{W_l}(\eta) \|_2 - \frac{M}{2^{l\beta_{max}}}$$

But by assumption of the theorem, $\frac{3\rho_n}{4} > \sum_{l=J_0}^{j*} \tau_l$

$\Rightarrow \exists\ \tau_l,\ J_0 \leq l \leq j*$ s.t.

$$\left(\| \Pi_{W_l}(\eta) \|_2 - \frac{M}{2^{l\beta_{max}}} - \tau_l \right) > 0$$, which proves the assertion of the lemma ☐

Once we have this lemma, the control of type II error of the test is easy. We know that for the $\{\tau_l\}$ satisfying the assumption of the lemma, $\exists\ J_0 \leq l \leq j*$ such that $\| \Pi_{W_l}(\eta) - \frac{M}{2^{l\beta_{max}}} - \tau_l \geq 0$.

For any such l, we have, with probability at least $1-\frac{\varepsilon}{2}$

$$T_n(l) \geq \|\Pi_{W_l}(\eta)\|_2^2 - c(\varepsilon,\psi,M,B)\sqrt{\frac{2^{(l+j_0)/2}}{n^2} + \frac{2^{l/4}\|\Pi_{W_l}(\eta)\|_2^2}{n}}$$

$$\geq \|\Pi_{W_l}(\eta)\|_2^2 - c(\varepsilon,\psi,M,B)\frac{2^{(l+j_0)/4}}{n} - c(\varepsilon,\psi,M,B)\,2^{l/8}\frac{\|\Pi_{W_l}(\eta)\|_2}{\sqrt{n}}$$

$$= \|\Pi_{W_l}(\eta)\|_2\left(\|\Pi_{W_l}(\eta)\|_2 - c(\varepsilon,\psi,M,B)\frac{2^{l/8}}{\sqrt{n}}\right) - c(\varepsilon,\psi,M,B)\frac{2^{(l+j_0)/4}}{n}$$

$$\geq \left(\frac{M}{2^{l\beta_{max}}}+\tau_l\right)\left(\frac{M}{2^{l\beta_{max}}}+\tau_l - c(\varepsilon,\psi,M,B)\frac{2^{l/8}}{\sqrt{n}}\right) - c(\varepsilon,\psi,M,B)\frac{2^{(l+j_0)/4}}{n} \quad (*)$$

Now I need to find $\{\tau_l\}$ s.t.

(i) $(*) > \left(\frac{M}{2^{l\beta_{max}}} + c(\varepsilon,\psi,M,B)\frac{2^{(l+j_*)/8}}{\sqrt{n}}\right)^2$ and

(ii) $\rho_n \geq \left(4\frac{M}{\sqrt{1-2^{-2\beta_{min}}}}2^{-j_*\beta_{min}} + \frac{4}{3}\sum_{j_0\leq l\leq j_*}\tau_l\right)$ where $\rho_n \gg n^{-\frac{4\beta_{min}}{4\beta_{min}+1}}$

<u>Check</u>: $\tau_l = C^*(\psi,M,B,\varepsilon)\frac{2^{(l+j_*)/8}}{\sqrt{n}}$ works for suitable C^* and $2^{j^*}\sim n^{\frac{2}{4\beta_{min}+1}}$

This completes the proof of the theorem. ▨

Next time we will use this theorem to address issues regarding construction of honest adaptive confidence sets over $\bigcup_{[\beta_{min},\beta_{max}]}\mathcal{P}(\beta,M,B)$ where $\beta_{max} > 2\beta_{min}$.