Supplement to "Adaptive Confidence Bands for Nonparametric Regression Functions"

Abstract

This supplement contains the proofs of Theorem 2, Propositions 2 and 3, Lemma 1 and Eq.(14).

7 Proof of Theorem 2

Rather than prove Theorem 2 directly it is convenient to first prove an analogue of the Theorem in the context of multivariate Normal random vectors. This is done in section 7.1. The proof of Theorem 2 is then given in section 7.2

7.1 Confidence Bound For Multivariate Normal Vectors

In the first proposition let $X_i, i = 1, 2, \ldots, n$ be independent Normal random variables, $N(c_n\theta_i, 1)$. Let $X = (X_1, X_2, \ldots X_n)$. Let $\theta = (\theta_1, \theta_2, \theta_n)$. For θ given we shall write $P^{X|\theta}$ and $E^{X|\theta}$ for computing probabilities and expectations under this model. We shall also assume that each θ_i is 0 or 1 and let Θ_n be the collection of such parameter values.

Suppose that $C(X) = (C_1(X), C_2(X), C_n(X))$ is a confidence set for $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ where $C_i(X)$ is a confidence interval for θ_i. Let $L(C_i(X))$ be the length of $C_i(X)$.

Proposition 4. *Suppose that $C(X)$ is a confidence set for θ with uniform coverage of at least $1 - \alpha$ over Θ_n. Suppose $c_n = \sqrt{c \log n}$ with $c < 1$. Then for any $a < 1$ and $\epsilon > 0$ there is an M such that for $n \geq M$*

$$\sup_{\theta \in \Theta_n} P^{X|\theta}\Big(\sum L(C_i(X)) \geq an\Big) \geq (1 - \alpha - \epsilon) \tag{89}$$

and hence for any $\epsilon > 0$

$$\sup_{\theta \in \Theta_n} E^{X|\theta}\Big(\sum L(C_i(X))\Big) \geq (1 - \alpha - \epsilon)n \tag{90}$$

when n is sufficiently large.

If the confidence set $C(X)$ also satisfies $L(C_i(X)) = L(C_1(X))$ for all i then for any $\epsilon > 0$ there is an M such that for $n \geq M$ and all $\theta \in \Theta_n$

$$E^{X|\theta}\Big(\sum L(C_i(X))\Big) \geq (1 - \alpha - \epsilon)n \tag{91}$$

For any $\alpha < \frac{1}{2}$ there is a $c > 0$ such that if $c_n = c$ then for any $\epsilon > 0$ there is an M and a $C > 0$ such that for $n \geq M$ and all $\theta \in \Theta_n$

$$E^{X|\theta}\Big(\sum L(C_i(X))\Big) \geq (1 - \alpha - \epsilon)Cn \tag{92}$$

Proof. First note that attention may be restricted to confidence bands where each $C_i(X)$ is equal either to the single points 0 or c_n or to the interval $[0, c_n]$. Put an equally likely independent prior on each coordinate and write π for this product prior. Write E for the expectation taken with respect to the joint distribution of the θ_i and the X_i and P for probabilities computed under such a model. Write E^π for the expected value with respect to the prior and write E^X for the expected value with respect to the marginal distribution of the vector X.

Now for any confidence band $C(X)$ let $C^l(X)$ be the band such that $C(X) = C^l(X)$ whenever $\sum_{i=1}^n L(C_i(X)) \leq lc_n$ and such that $C^l(X) = (0, 0, \ldots, 0)$ otherwise. For any $0 < a < 1$ let $B_n(a)$ be the event that $\sum_{i=1}^n L(C_i(X)) \leq anc_n$. Then

$$E(1(\theta \in C(X))) \leq E(1(\theta \in C(X))1(B_n(a)))+(1-P(B_n(a))) \leq E(1(\theta \in C^{an}(X)))+(1-P(B_n(a))) \tag{93}$$

Let

$$N = \sum_{i=1}^n (1(-f \leq X_i \leq f) + 1(c_n - f \leq X_i \leq c_n + f)) \tag{94}$$

For any $a < d < 1$, let $A_n(d)$ be the event that $N \geq dn$. Note that the marginal distribution of N is binomial. Let f be any value such that for any $\epsilon > 0$ if $n > M$, $P(A_n(d)) \geq 1 - \epsilon$. Write ϕ for the density function of a standard Normal random variable. Since $\phi(y) \leq \frac{1}{2}$ it follows that $P(\theta_i = c_n|X_i) = \frac{\frac{1}{2}\phi(X_i - c_n)}{\frac{1}{2}\phi(X_i - c_n) + \frac{1}{2}\phi(X_i)} \geq \phi(X_i - c_n)$ and hence for each X_i in the interval $[-f, f]$, $P(\theta_i = c_n|X_i) \geq \phi(f + c_n)$. Similarly For each X_i in $[c_n - f, c_n + f]$ we also have $P(\theta_i = 0|X_i) \geq \phi(f + c_n)$.

Now let $p_i = \min(P(\theta_i = 0|X_i), P(\theta_i = c_n|X_i))$ and let $p_{(j)}(X)$ be the jth smallest of these.

$$E(1(\theta \in C^{an}(X))) = E^\pi P^{X|\theta}(\theta \in C^{an}(X)) \leq E^X \prod_{j=1}^{n-an} (1 - p_{(j)}(X)) \tag{95}$$

Also if $n > M$

$$E^X \prod_{j=1}^{n-an} (1-p_{(j)}(X)) \leq E^X(\prod_{j=1}^{n-an} (1-p_{(j)}(X))|A_n(d))+(1-P(A_n(d)) \leq (1-\phi(f+c_n))^{dn-an}+\epsilon \tag{96}$$

Hence

$$E(1(\theta \in C(X))) \leq (1 - \phi(f + c_n))^{dn-an} + \epsilon + (1 - P(B_n(a))) \tag{97}$$

Since $c < 2$ it also follows that for sufficiently large n

$$(1 - \phi(f + c_n))^{dn-an} < \epsilon \tag{98}$$

Hence for sufficiently large n

$$E(1(\theta \in C(X))) \leq 2\epsilon + (1 - P(B_n(a))) \tag{99}$$

Now if for all $\theta \in \Theta_n$

$$P^{X|\theta}(\theta \in C(X)) \geq 1 - \alpha \tag{100}$$

it follows that for sufficiently large n

$$1 - \alpha \leq E(1(\theta \in C(X))) \leq 2\epsilon + (1 - P(B_n(a))) \tag{101}$$

and hence

$$P(B_n(a)) \leq \alpha + 2\epsilon \tag{102}$$

It then follows that there is a $\theta \in \Theta_n$ such that

$$P^{X|\theta}(B_n(a)) \leq \alpha + 2\epsilon \tag{103}$$

For this θ it follows that for sufficiently large n

$$P^{X|\theta}(\sum_{i=1}^{n} L(C_i(X)) \geq an) \geq 1 - \alpha - 2\epsilon. \tag{104}$$

Since ϵ is arbitrary (89) follows immediately. From (89) it is easy to see that (90) follows immediately.

We now turn to the proof of (91) which without loss of generality we shall prove for $\theta = \theta^0 = (0, 0, \ldots 0)$ the zero vector. For this we need to do mixing. Let $\theta^i = (0, 0, \ldots, 1, 0, \ldots 0)$. where 1 occurs in place i. Let $Q = \frac{1}{n}\sum_{i=1}^{n} P^{X|\theta^i}$. Denote by ψ_i the density of $P^{X|\theta^i}$ for $i = 0, 1, ..., n$, then straightforward calculations yield that for $1 \leq i, j \leq n$

$$\int \frac{\psi_i \psi_j}{\psi_0} = \exp(\delta_{ij} c \log n)$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. Then the chi-squared distance between $P^{X|\theta^0}$ and the mixture Q satisfies

$$\chi^2(P^{X|\theta^0}, Q) = E^{X|\theta^0}\left(\frac{\frac{1}{n}\sum_i^n \psi_i}{\psi_0} - 1\right)^2 = \frac{1}{n^2}\sum_{i,j=1}^{n}\int \frac{\psi_i\psi_j}{\psi_0} - 1 = \frac{1}{n}(\exp(c\log n) - 1) \leq n^{-(1-c)}.$$

In particular, $\chi(P^{X|\theta^0}, Q) \to 0$ for any constant $0 < c < 1$. Consequently, the Total Variation distance between $P^{X|\theta^0}$ and Q satisfies

$$||P^{X|\theta^0} - Q||_{TV} = \frac{1}{2}\int\left|\psi_0 - \frac{1}{n}\sum_i^n \psi_i\right| \leq \frac{1}{2}\chi(P^{X|\theta^0}, Q) \to 0. \tag{105}$$

When all the confidence intervals are of the same length $\theta^0 \in C(X)$ when $C(X) = \theta^0$ or when $C(X) = ([0, 1], [0, 1], \ldots, [0, 1])$. Hence

$$P^{X|\theta^0}(\theta^0 \in C(X)) = P^{X|\theta^0}(C(X) = \theta^0) + P^{X|\theta^0}(C(X) = ([0, 1], [0, 1], \ldots, [0, 1])) \tag{106}$$

For any $\epsilon > 0$, from (105) it follows that there is an M such that for $n > M$

$$||P^{X|\theta^0} - Q||_{TV} \leq \epsilon \tag{107}$$

Note that since $C(X)$ is assumed to have coverage probability of at least $1-\alpha$ it follows that for $i = 1, 2, \ldots n$

$$P^{X|\theta^i}(C(X) = \theta^0) < \alpha \tag{108}$$

and so

$$Q(C(X) = \theta^0) < \alpha. \tag{109}$$

Hence if $n > M$ from (107) it follows that $P^{X|\theta^0}(C(X) = \theta^0) \leq \alpha + \epsilon$. Hence for $n > M$, $P^{X|\theta^0}(C(X) = ([0,1], [0,1], \ldots, [0,1])) \geq 1 - 2\alpha - \epsilon$ and it follows that

$$E^{X|\theta^0}(\sum L(C_i(X)) \geq (1 - 2\alpha - \epsilon)n \tag{110}$$

which proves (91).

In the proof of (92) we may without loss of generality also take $\theta^0 = (0, 0, \ldots, 0)$. For any $\alpha < 1$ choose c so that for each $i = 1, \ldots n$, $||P_{\theta^0} - P_{\theta^i}||_{TV} \leq \epsilon$ whenever $n > M$. Then $P_{\theta^i}(C_i(X) < \alpha)$ and hence $P_{\theta^0}(C_i(X) = 0) \leq \alpha + \epsilon$. Hence $P_{\theta^0}(C_i(X) = [0,1]) \geq 1 - 2\alpha - \epsilon$. and it follows that there is a $C > 0$ such that for each i, $EL(C_i(X)) \geq C$ which immediately yields (92).

7.2 Proof of Theorem 2

Without loss of generality we shall assume the noise level $\sigma = 1$. Let g be an infinitely differentiable function supported on $[0,1]$ with $g(t) > 0$ for $t \in (0,1)$ and $\int_0^1 g^2(t)dt = 1$. For instance, one can set

$$g(t) = \begin{cases} c_g \left(\exp\left(-\frac{1}{t} e^{-\frac{1}{1-t}}\right) + \exp\left(-\frac{1}{1-t} e^{-\frac{1}{t}}\right) - 1 \right), & t \in [0,1], \\ 0, \quad \text{otherwise.} \end{cases} \tag{111}$$

Here, the normalizing constant $c_g \doteq 0.346$.

Let m be a positive integer and set $k_n = \frac{n}{m}$. For ease of exposition we shall assume n is divisible by m and so k_n is an integer but all the arguments that follow hold under obvious modifications if we take k_n to be the integer part of $\frac{n}{m}$. Set $A_{k_n} = \frac{1}{k_n}\sum_{j=1}^{k_n} g^2(\frac{j}{k_n})$ and note that A_{k_n} is bounded since g is bounded. Fix an $f \in \Lambda(\beta, M')$ with $M' < M$. Then for $\theta = (\theta_1, \theta_2, \ldots \theta_m)$, define the function f_θ by

$$f_\theta(t) = f(t) + \frac{1}{A_{k_n}^{\frac{1}{2}}} \sum_{i=1}^{m} \theta_i c_0 (M - M') m^{-\beta} g(m(t - x_i)) \tag{112}$$

where $x_i = \frac{i-1}{m}$ and $c_0 > 0$ is a constant. Since A_{k_n} is bounded it is easy to verify that, when the constant c_0 is chosen sufficiently small, all realizations of f are in $\Lambda(\beta, M)$. Let

$$Y_i = (\frac{1}{A_{k_n} k_n})^{1/2} \sum_{j=1}^{k_n} g(\frac{j}{k_n})(y_{(i-1)k_n+j} - f(x_i + \frac{j}{n})) \tag{113}$$

Note that $Y = (Y_1, Y_2, \dots Y_m)$ is sufficient for θ. The Y_i are independent and Normal with

$$E(Y_i) = (\frac{1}{k_n})^{1/2} \frac{1}{A_{k_n}} \theta_i \sum_{j=1}^{k_n} c_0(M - M')m^{-\beta} g^2(\frac{j}{k_n}) = c_0(M - M')\theta_i n^{\frac{1}{2}} m^{\frac{-(2\beta+1)}{2}} \tag{114}$$

and

$$Var(Y_i) = \frac{1}{A_{k_n} k_n} \sum_{j=1}^{k_n} g^2(\frac{j}{k_n}) = 1 \tag{115}$$

We now prove (45) and (46). Here we take $f(t) \equiv 0$ in (112) in which case $M' = 0$. For confidence intervals of $f_\theta(t)$ over the class Θ_n we may restrict attention to confidence bands $CB(t) = \sum_{i=1}^m C_i(Y) \frac{1}{A_{k_n}^{\frac{1}{2}}} c_0 M m^{-\beta} g(m(t - x_i))$ where $C_i(Y)$ is a confidence interval for θ_i.

Let $\gamma = \int_0^1 g(t)dt$. Then

$$\int_0^1 m^{-\beta} g(m(t - x_i))dt = m^{-(\beta+1)}\gamma. \tag{116}$$

Hence

$$\int_0^1 CB(t) = \frac{1}{A_{k_n}^{\frac{1}{2}}} c_0 M m^{-(\beta+1)} \gamma \sum_{i=1}^m L(C_i(Y)) \tag{117}$$

Now set $m = \lceil (\frac{M^2 n}{\log n})^{\frac{1}{2\beta+1}} \rceil$. Then

$$E(Y_i) = c_0 \theta_i \sqrt{\log n}. \tag{118}$$

Since c_0 can be selected to be any positive real number take $c_0 < 1$ such that f_θ in (112) is guaranteed to be in $\Lambda(\beta, M)$. It follows from (89) that for any $a < 1$ there is an N such that for $n \geq N$

$$\sup_{\theta \in \Theta_n} P(\int_0^1 CB(t) \geq \frac{1}{A_{k_n}^{\frac{1}{2}}} c_0 M m^{-(\beta+1)} \gamma a m) \geq 1 - \alpha - \epsilon. \tag{119}$$

and from (90) that

$$\sup_{\theta \in \Theta_n} E(\int_0^1 CB(t)) \geq C \frac{1}{A_{k_n}^{\frac{1}{2}}} c_0 M m^{-(\beta+1)} \gamma a m. \tag{120}$$

Since $Mm^{-\beta} = (\frac{\log n}{n})^{\frac{\beta}{2\beta+1}}$ equation (45) and (46) immediately follow.

We now turn to the proof of equation (48). Fix an $f \in \Lambda(\beta, M')$ with $M' < M$. Take $m = \lceil (\frac{(M-M')^2 n}{\log n})^{\frac{1}{2\beta+1}} \rceil$. Once again

$$E(Y_i) = c_0 \theta_i \sqrt{\log n} \tag{121}$$

and (48) follows from (91).

Finally take $m = \lceil ((M-M')^2 n)^{\frac{1}{2\beta+1}} \rceil$ and in this case

$$E(Y_i) = c_0 \theta_i \tag{122}$$

and by taking a sufficiently small c_0 (47) follows from (92).

8 Proofs of Propositions 2 and 3

This section is dedicated to the proofs of Propositions 2 and 3. To this end, we first investigate a general form of the levelwise test in Section 8.1, followed by the proofs of both propositions in Section 8.2.

8.1 General Levelwise Tests

After proper scaling (by a factor of $\sigma_n = \sigma n^{-\frac{1}{2}}$), all the level-wise hypotheses $H_{0,jl}$ defined in (31) share the same form as the following. Let $X_i \overset{ind}{\sim} N(\theta_i, 1)$, for $i = 1, \ldots, m$. We want to test

$$H_0 : \max_{1 \le i \le m} |\theta_i| \le c_m. \tag{123}$$

Here, $n > m = 2^l$ for some integer $l < J$, and c_m is identified with some $\sigma_n^{-1} c_{jl}$. Depending on (j, l), c_m could range from $O((\log m)^{\frac{1}{2}})$ to $O(m^{-q})$ for some $q > 0$. On the other hand, the test statistics that we use, after proper scaling, becomes $\max_i |X_i|$ and

$$T_m(t_m) = \sum_{i=1}^{m} |X_i| I_{\{|X_i| > t_m\}}, \tag{124}$$

where t_m depends on c_m. Indeed, after proper scaling, the events in (34) become

$$\mathcal{R}_0 = \left\{ \max_i |X_i| > (\sqrt{3} + \sqrt{2}) \sqrt{\log n} \right\}, \tag{125}$$

$$\mathcal{R}_1 = \left\{ T_m(t_m) > m\mu(c_m; t_m) + \tfrac{1}{2}(m \log n)^{\frac{1}{2}} [c_m + \left(\tfrac{5}{2} \log n\right)^{\frac{1}{2}}] \right\}, \quad \text{and} \tag{126}$$

$$\mathcal{R}_2 = \left\{ T_m(1) > m\mu(c_m; 1) + [(1 + c_m^2) m \log n]^{\frac{1}{2}} \right\}, \tag{127}$$

where μ is defined in (33) and $t_m = c_m + (2r \log n)^{\frac{1}{2}}$. The choice made after (32) corresponds to $r = \log_2 m / (4 \log n)$, which guarantees that $r < 1/4$ for all the relevant values of m when n is sufficiently large.

8.1.1 Control of Type I Errors

We first investigate the type I error of the rejection regions $\mathcal{R}_0 \cup \mathcal{R}_1$ and $\mathcal{R}_0 \cup \mathcal{R}_2$. To this end, the following lemma is helpful.

Lemma 2. *Suppose $X \sim N(\theta, 1)$, then*

$$\mathsf{E}_\theta |X| I_{\{|X|>t\}} = \mu(\theta; t) = \phi(t+\theta) + \phi(t-\theta) + \theta[\Phi(t+\theta) - \Phi(t-\theta)],$$
$$\mathsf{E}_\theta X^2 I_{\{|X|>t\}} = (t+\theta)\phi(t-\theta) + (t-\theta)\phi(t+\theta) + (\theta^2+1)[2 - \Phi(t-\theta) - \Phi(t+\theta)].$$

In addition, $\mu'(\theta; t) = \frac{d}{d\theta}\mu(\theta; t) = t[\phi(t-\theta) - \phi(t+\theta)] + \Phi(t+\theta) - \Phi(t-\theta)$.

Proof. By symmetry, we consider only the case where $\theta \geq 0$.

For the first moment, we have

$$\mathsf{E}_\theta |X| I_{\{|X|>t\}} = \int_t^\infty x\phi(x-\theta)dx + \int_{-\infty}^{-t} (-x)\phi(x-\theta)dx.$$

For the first term, we have

$$\int_t^\infty x\phi(x-\theta)dx = \int_{t-\theta}^\infty y\phi(y)dy + \theta\int_{t-\theta}^\infty \phi(y)dy = \phi(t-\theta) + \theta[1 - \Phi(t-\theta)].$$

Note that the equality holds regardless of whether $t \geq \theta$. Similarly, we obtain $\int_{-\infty}^{-t}(-x)\phi(x-\theta)dx = \phi(t+\theta) - \theta[1-\Phi(t+\theta)]$. Putting the two parts together leads to the claimed formula.

Turn to the second moment. We have

$$\mathsf{E}_\theta X^2 I_{\{|X|>t\}} = \int_t^\infty x^2\phi(x-\theta)dx + \int_{-\infty}^{-t} x^2\phi(x-\theta)dx.$$

Focus on the first term on the right side. We have

$$\begin{aligned}\int_t^\infty x^2\phi(x-\theta)dx &= \int_t^\infty (x-\theta)^2\phi(x-\theta)dx + 2\theta\int_t^\infty (x-\theta)\phi(x-\theta)dx + \theta^2\int_t^\infty \phi(x-\theta)dx \\ &= (t-\theta)\phi(t-\theta) + [1-\Phi(t-\theta)] + 2\theta\phi(t-\theta) + \theta^2[1-\Phi(t-\theta)] \\ &= (t+\theta)\phi(t-\theta) + (\theta^2+1)[1-\Phi(t-\theta)].\end{aligned}$$

Here, the second equality uses the identity $\int_t^\infty x^2\phi(x)dx = t\phi(t) + 1 - \Phi(t)$. Again, the above equalities hold regardless of whether $t \geq \theta$. By symmetry, we obtain an analogous expression for the second term in the second last display. Combining the two parts, we obtain the formula.

Finally, the last formula is obtained by directly differentiating the first expression with respect to θ. This completes the proof. ■

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)' \in \mathbb{R}^m$. For the rejection region $\mathcal{R}_0 \cup \mathcal{R}_1$, we have the following result.

Lemma 3. *There exists a constant C, such that for all m with $c_m \le \sqrt{2\log n}$, and all $\boldsymbol{\theta} \in H_0$, $P_{\boldsymbol{\theta}}(\mathcal{R}_0 \cup \mathcal{R}_1) \le Cn^{-\frac{1}{2}}$.*

Proof. Recall that $t_m = c_m + (2r\log m)^{\frac{1}{2}}$, $r < \frac{1}{2}$ and $n > m$. Let $b_m = (\frac{5}{2}\log n)^{\frac{1}{2}} > (2r\log m)^{\frac{1}{2}}$. Define $U_i = |X_i| I_{\{|X_i|>t_m, |X_i|\le c_m+b_m\}}$. For all $\boldsymbol{\theta} \in H_0$ and any $x > 0$,

$$\begin{aligned} &P_{\boldsymbol{\theta}}\left(T_m(t_m) - m\mu(c_m;t_m) > x\right) \\ &\quad \le P_{\boldsymbol{\theta}}\left(\sum_{i=1}^m U_i - m\mu(c_m;t_m) > x\right) + P_{\boldsymbol{\theta}}\left(\exists i, |X_i| > c_m + b_m\right). \end{aligned} \tag{128}$$

Under H_0, we have $U_i \in [0, c_m + b_m]$ and $\mathsf{E}U_i \le \mu(c_m;t_m)$. So, Hoeffding's inequality leads to

$$P_{\boldsymbol{\theta}}\left(\sum_{i=1}^m U_i - m\mu(c_m;t_m) > x\right) \le \exp\left\{-\frac{2x^2}{m(c_m+b_m)^2}\right\}.$$

In addition, a simple union bound leads to

$$P_{\boldsymbol{\theta}}\left(\exists i, |X_i| > c_m + b_m\right) \le \sum_{i=1}^m P_{\theta_i}(|X_i| > c_m + b_m) \le mP(|Z| > b_m) \le 2m\exp\left(-\frac{b_m^2}{2}\right).$$

Here, Z stands for a standard normal random variable. We obtain $P_{\boldsymbol{\theta}}(\mathcal{R}_1) \le Cn^{-\frac{1}{2}}$ by fixing $x = \frac{1}{2}(m\log n)^{\frac{1}{2}}[c_m + (\frac{5}{2}\log n)^{\frac{1}{2}}]$ and bounding the two terms on the right side of (128) by the last two displays.

In addition, since $c_m \le \sqrt{2\log n}$, for any $\boldsymbol{\theta} \in H_0$, we have

$$P_{\boldsymbol{\theta}}(\mathcal{R}_0) \le \sum_{i=1}^m P_{\theta_i}(|X_i| > (\sqrt{3}+\sqrt{2})\sqrt{\log n}) \le mP(|Z| > \sqrt{3\log n}) \le Cn^{-\frac{1}{2}}.$$

We completes the proof by noting that $P_{\boldsymbol{\theta}}(\mathcal{R}_0 \cup \mathcal{R}_1) \le P_{\boldsymbol{\theta}}(\mathcal{R}_0) + P_{\boldsymbol{\theta}}(\mathcal{R}_1)$. ∎

Turning to the rejection region $\mathcal{R}_0 \cup \mathcal{R}_2$, we focus on the cases where $c_m \le (\log n)^{-\frac{1}{2}}$.

Lemma 4. *There exists a constant C, such that for all m with $c_m \le (\log n)^{-\frac{1}{2}}$, and all $\boldsymbol{\theta} \in H_0$, $P_{\boldsymbol{\theta}}(\mathcal{R}_0 \cup \mathcal{R}_2) \le Cn^{-\frac{1}{2}}$.*

Proof. We first show that $P_{\boldsymbol{\theta}}(\mathcal{R}_2) \le Cn^{-\frac{1}{2}}$. Let $Y_i = |X_i| I_{\{|X_i|>1\}} - \mathsf{E}|X_i| I_{\{|X_i|>1\}}$, then $\mathsf{E}Y_i = 0$. When $|\theta_i| \le c_m$, $\mathsf{E}Y_i^2 \le \mathsf{E}X_i^2 = 1 + \theta_i^2 \le 1 + c_m^2$. Moreover, for any integer $p \ge 2$,

$$\mathsf{E}|Y_i|^p \le 2^{p-1}[\mathsf{E}|X_i|^p I_{\{|X_i|>1\}} + (\mathsf{E}|X_i| I_{\{|X_i|>1\}})^p] \le 2^{m-1}[\mathsf{E}|X_i|^p + (\mathsf{E}|X_i|)^p].$$

Note that $\mathsf{E}|X_i| \le \sqrt{2/\pi} + c_m$, and that

$$\mathsf{E}|X_i|^p \le 2^{p-1}\left(\mathsf{E}|X_i - \theta_i|^p + |\theta_i|^p\right) \le 2^{p-1}[(p-1)!! + c_m^p].$$

We thus have for all $p \geq 2$,

$$\mathsf{E}|Y_i|^p \leq 4^{p-1}(p-1)!! + 4^{p-1}c_m^p + 2^{p-1}[\sqrt{2/\pi} + c_m]^p \leq \tfrac{1}{2}p!\, 3^{p-2}.$$

In the last inequality, we use the assumption that $c_m \leq (\log n)^{-\frac{1}{2}}$. Then for any $x > 0$ and all $\boldsymbol{\theta} \in H_0$, Bernstein's inequality leads to

$$P_{\boldsymbol{\theta}}\left(T_m(1) > \mathsf{E}_{\boldsymbol{\theta}} T_m(1) + x\right) \leq \exp\left\{-\frac{1}{2}\frac{x^2}{(1+c_m^2)m + 3x}\right\}.$$

We complete the part by fixing $x = [(1+c_m^2)m\log n]^{\frac{1}{2}}$ and noting that $\mathsf{E}_{\boldsymbol{\theta}} T_m(1) \leq m\mu(c_m;1)$ for all $\boldsymbol{\theta} \in H_0$.

On the other hand, we repeat as in the proof of Lemma 3 to conclude that $P_{\boldsymbol{\theta}}(\mathcal{R}_0) \leq Cn^{-\frac{1}{2}}$. This completes the proof. ■

8.1.2 Power of Tests

We now derive finite sample lower bounds for the power of the rejection regions $\mathcal{R}_0 \cup \mathcal{R}_1$ and $\mathcal{R}_0 \cup \mathcal{R}_2$ against two different types of alternative hypotheses: the excess mass type and the noncovered points type.

Excess mass type alternative In this case, we are interested in testing H_0 against

$$H_1 : \sum_{i=1}^{m} \left(|\theta_i| - c_m\right)_+ > e_m. \tag{129}$$

Lemma 5. *Consider testing H_0 (123) against H_1 (129) based on the rejection region $\mathcal{R}_0 \cup \mathcal{R}_1$. Suppose that $(\log n)^{-\frac{1}{2}} \leq c_m \leq C_1(\log n)^{\frac{1}{2}}$ and that $C_2 \log m \geq \log n$. In t_m, let $r < \frac{1}{2}$ be a fixed constant. If $e_m = Cm(\log n)^{-\frac{1}{2}}$ for a sufficiently large constant C, then the power of $\mathcal{R}_0 \cup \mathcal{R}_1$ is at least $1 - C_3 \log n/m^{1-2r}$.*

Proof. Recall that $t_m = c_m + (2r\log m)^{\frac{1}{2}}$. Let $d_m = \exp(-c_m(t_m - \frac{1}{2}c_m))$. We have

$$\mu(0;t_m) = \mathsf{E}_0|X|I_{\{|X|>t_m\}} = 2\phi(t_m) = 2(2\pi)^{-\frac{1}{2}}m^{-r}d_m.$$

In addition, write $a_m = \exp(-2c_m t_m)$, Lemma 2 leads to

$$\mu'(c_m;t_m) \geq (2\pi)^{-\frac{1}{2}}m^{-r}t_m(1-a_m).$$

Since $\mu(\theta;t_m)$ is increasing in $|\theta|$ when $|\theta| \leq c_m$, we have

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(t_m) - m\mu(c_m;t_m) \geq \sum_{|\theta_i|>c_m} [\mu(|\theta_i|;t_m) - \mu(c_m;t_m)] - m[\mu(c_N;t_m) - \mu(0;t_m)].$$

Note that $\mu'(c_m; t_m) = \inf_{|\theta| \geq c_m} \mu'(\theta; t_m) > 0$, thus the first term on the right side is bounded below by $\mu'(c_m; t_m) \sum_i (|\theta_i| - c_m)_+ \geq \mu'(c_m; t_m) e_m$. Together with the last three displays, this leads to

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(t_m) - m\mu(c_m; t_m) \geq (2\pi)^{-\frac{1}{2}} m^{-r} \left[t_m(1 - a_m) e_m - m \left(\frac{t_m}{t_m - c_m} + \frac{a_m t_m}{t_m + c_m} - 2d_m \right) \right].$$

Since $c_m \geq (\log n)^{-\frac{1}{2}}$, $c_m t_m$ is bounded away from 0, and so $1 - a_m \geq C_4$ for some positive constant C_4. In addition, $c_m \leq C_1 (\log n)^{\frac{1}{2}}$ implies $t_m/(t_m - c_m) \leq C_5$. Thus, for $e_m = Cm(\log n)^{-\frac{1}{2}}$ with a sufficiently large C, we obtain

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(t_m) - m\mu(c_m; t_m) \geq C_6 (C - C_7) m^{1-r} \geq 2B_{1m}, \tag{130}$$

where $B_{1,m} = \frac{1}{2}(m \log n)^{\frac{1}{2}} [c_m + \left(\frac{5}{2} \log n\right)^{\frac{1}{2}}]$.

Moreover, $c_m \leq C_1 (\log n)^{\frac{1}{2}}$ implies that $t_m \leq C_8 (\log n)^{\frac{1}{2}}$. This, together with Lemma 2, implies that $\mathsf{Var}_\theta(|X| I_{\{|X| > t_m\}}) \leq C_9 \log n$ for any θ. So, for any $\boldsymbol{\theta} \in H_1$,

$$\mathsf{Var}_{\boldsymbol{\theta}}(T_m(t_m)) \leq C_{10} m \log n. \tag{131}$$

Thus, for any $\boldsymbol{\theta} \in H_1$, the type II error of $\mathcal{R}_0 \cup \mathcal{R}_1$ is

$$P_{\boldsymbol{\theta}}((\mathcal{R}_0 \cup \mathcal{R}_1)^c) \leq P_{\boldsymbol{\theta}}(\mathcal{R}_1^c) \leq P_{\boldsymbol{\theta}}(T_m(t_m) - \mathsf{E}_{\boldsymbol{\theta}} T_m(t_m) \leq -C_6 (K - C_7) m^{1-r} + B_{1m}).$$

To bound the right side, we apply Chebyshev's inequality, together with (130) and (131), to obtain $P_{\boldsymbol{\theta}}(\mathcal{R}_1^c) \leq \mathsf{Var}_{\boldsymbol{\theta}}(T_m(t_m))/[C_6(K - C_7) m^{1-r} - B_{1m}]^2 \leq C_3 \log n / m^{1-2r}$. ■

Lemma 6. *Consider testing H_0 (123) against H_1 (129) based on the rejection region $\mathcal{R}_0 \cup \mathcal{R}_2$. Suppose that $C_1 (\log m/m)^{1/4} \leq c_m \leq (\log n)^{-\frac{1}{2}}$ and that $C_2 \log m \geq \log n$. If $e_m = Cmc_m$ for a sufficiently large constant C, then the power of $\mathcal{R}_0 \cup \mathcal{R}_2$ is at least $1 - C_3 n^{-\frac{1}{2}}$.*

Proof. Similar to the previous case, for any $\boldsymbol{\theta} \in H_1$, we have

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(1) - m\mu(c_m; 1) \geq \sum_{|\theta_i| > c_m} [\mu(|\theta_i|; 1) - \mu(c_m; 1)] - m[\mu(c_m; 1) - \mu(0; 1)].$$

Lemma 2 leads to $\mu'(c_m; 1) = \inf_{\theta > c_m} \mu'(\theta; 1) = \sup_{\theta \in [0, c_m]} \mu'(\theta; 1)$ when $c_m \leq 1/2$. Thus, we could further bound the right side of the last display and obtain

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(1) - m\mu(c_m; 1) \geq \mu'(c_m; 1)(e_m - mc_m).$$

To further control the right side, note that Lemma 2 leads to $\mu'(c_m; 1) \geq 2c_m \phi(1 + c_m) \geq C_4 c_m$, and so for all $\boldsymbol{\theta} \in H_1$,

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(1) - m\mu(c_m; 1) \geq C_4 (C - 1) m c_m^2. \tag{132}$$

In addition, Lemma 2 implies that for all $\boldsymbol{\theta} \in H_1$, $C_5 m \le \mathsf{Var}_{\boldsymbol{\theta}}(T_m(1)) \le C_6 m$.

To obtain the desired lower bound for power, we divide into two cases.

First, if $\max_i |\theta_i| \le C_7 (\log n)^{\frac{1}{2}}$, Lemma 2 implies that there exists $t'_m = C_8 (\log n)^{\frac{1}{2}}$, such that

$$\mathsf{E}_{\theta_i} |X_i| I_{\{|X_i| > t'_m\}},\ \mathsf{E}_{\theta_i} X_i^2 I_{\{|X_i| > t'_m\}},\ P_{\boldsymbol{\theta}}(\max_i |\theta_i| > t'_m) \le n^{-2}. \tag{133}$$

Now define

$$T'_m(1) = T_m(1) - T_m(t'_m) = \sum |X_i| I_{\{1 < |X_i| \le t'_m\}}. \tag{134}$$

Then (133) leads to

$$\mathsf{E}_{\boldsymbol{\theta}} T'_m(1) - m\mu(c_m; 1) \ge C'_4 (C-1) m c_m^2, \quad C'_5 m \le \mathsf{Var}_{\boldsymbol{\theta}}(T'_m(1)) \le C'_6 m. \tag{135}$$

Let $B_{2m} = [(1 + c_m^2) m \log n]^{\frac{1}{2}}$, then $C'_4(C-1) m c_m^2 > 2B_{2m}$ for sufficiently large C. Thus, we can bound the probability of type II error of $\mathcal{R}_2$ as

$$P_{\boldsymbol{\theta}}(\mathcal{R}_2^c) \le P_{\boldsymbol{\theta}}(T'_m(1) - m\mu(c_m; 1) \le B_{2m}) + P_\theta(T_m(1) \ne T'_m(1)).$$

For the first term, (135) and the discussion after it imply that it is bounded by

$$P_{\boldsymbol{\theta}}(T'_m(1) - E_{\boldsymbol{\theta}} T'_m(1) \le -\tfrac{1}{2} C'_4 (C-1) m c_m^2).$$

Note that each summand in $T'_m(1)$ is bounded by t'_m, and so we could apply Bennett's inequality [this version due to Devroye and Lugosi (2001)] to bound the probability in the last display by

$$\exp\left\{ -\frac{\mathsf{Var}_{\boldsymbol{\theta}}(T'_m(1))}{(t'_m)^2} g\left(\frac{t'_m C'_5 (C-1) m c_m^2}{2\mathsf{Var}_{\boldsymbol{\theta}}(T'_m(1))} \right) \right\},$$

where $g(x) = (1+x)\log(1+x) - x \ge x^2/4$ for $x \in (0, 1/2)$. Based on the above discussion, the argument of g is of order $O(m^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) = o(1)$. Thus, for a sufficiently large C,

$$P_{\boldsymbol{\theta}}(T'_m(1) - E_{\boldsymbol{\theta}} T'_m(1) \le -\tfrac{1}{2} C'_4 (C-1) m c_m^2) \le \exp\left\{ -C_8 (C-1)^2 m c_m^4 \right\} \le (C_3/2) n^{-\frac{1}{2}}.$$

Further note that $P_\theta(T_m(1) \ne T'_m(1)) \le P_{\boldsymbol{\theta}}(\max_i |\theta_i| > t'_m) \le n^{-2}$. The triangle inequality thus leads to $P_{\boldsymbol{\theta}}(\mathcal{R}_2^c) \le C_3 n^{-\frac{1}{2}}$.

Next, if $\max_i |\theta_i| > C_7 (\log n)^{\frac{1}{2}}$ for a large enough constant $C_7 > 2\sqrt{2} + \sqrt{3}$, then for $Z \sim N(0,1)$, $P_{\boldsymbol{\theta}}(\mathcal{R}_0^c) \le m P(|Z| > \sqrt{3 \log n}) \le C_3 n^{-\frac{1}{2}}$. This completes the proof. ■

Lemma 7. *Consider testing* H_0 (123) *against* H_1 (129) *based on the rejection region* $\mathcal{R}_0 \cup \mathcal{R}_2$. *Suppose that* $0 < c_m \le C_1 (\log m/m)^{1/4}$ *and that* $C_2 \log m \ge \log n$. *If* $e_m = C m^{\frac{3}{4}} (\log m)^{\frac{1}{4}}$ *for a sufficiently large constant* C, *then the power of* $\mathcal{R}_0 \cup \mathcal{R}_2$ *is at least* $1 - C_3 n^{-\frac{1}{2}}$.

Proof. Let $\eta_m = \frac{C}{2}(\log m/m)^{\frac{1}{4}} > 2c_m$ for sufficiently large C. Since $e_m = Cm^{\frac{3}{4}}(\log m)^{\frac{1}{4}}$, we must have $\sum_{|\theta_i|>\eta_m}(|\theta_i| - c_m)_+ \geq e_m/2$. For any $\boldsymbol{\theta} \in H_1$, this leads to the bound

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(1) - m\mu(c_m; 1) \geq \sum_{|\theta_i|>\eta_m} [\mu(|\theta_i|; 1) - \mu(c_m; 1)] - m[\mu(c_m; 1) - \mu(0; 1)].$$

Note that $\mu(\theta; 1)$ is convex on $[0, 1]$, so for any $|\theta_i| > \eta_m > 2c_m$,

$$\mu(|\theta_i|; 1) - \mu(c_m; 1) \geq \mu'(\eta_m/2; 1)(|\theta_i| - c_m).$$

Further note that $\mu'(c_m; 1) = \sup_{\theta \in [0,c_m]} \mu'(\theta; 1)$, we thus obtain

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(1) - m\mu(c_m; 1) \geq \mu'(\eta_m/2; 1)e_m/2 - mc_m\mu'(c_m; 1).$$

Lemma 2 implies that for some constant $C_4 > 0$, $\mu'(\eta_m/2; 1) > C_4\eta_m$ and $\mu'(c_m; 1) \geq C_4c_m$. This, together with the last display, implies

$$\mathsf{E}_{\boldsymbol{\theta}} T_m(1) - m\mu(c_m; 1) \geq C_5(C^2 - 1)(m \log n)^{\frac{1}{2}}.$$

Moreover, Lemma 2 also implies that for all $\boldsymbol{\theta} \in H_1$, $C_6 m \leq \mathsf{Var}_{\boldsymbol{\theta}}(T_m(1)) \leq C_7 m$. The proof could then be completed by repeating the two case argument in the proof of Lemma 6. ■

Noncovered points type alternative To deal with the set of noncovered points, we test (123) against

$$H_1' : m^{-1}|\{\theta_i : |\theta_i| > \tilde{c}_m\}| > \kappa_m. \tag{136}$$

For each m, $\tilde{c}_m > c_m$ depends on the value of c_m and is to be specified below.

Lemma 8. *Consider testing H_0 (123) against H_1' (136) based on the rejection region $\mathcal{R}_0 \cup \mathcal{R}_1$. Suppose that $(\log n)^{-\frac{1}{2}} \leq c_m \leq C_1(\log n)^{\frac{1}{2}}$ and that $C_2 \log m \geq \log n$. In t_m, let $r < \frac{1}{2}$ be a fixed constant. If in (136), $\tilde{c}_m = (\gamma_m + 1)c_m$ for some $\gamma_m > 0$ and $\kappa_m = C(\gamma_m c_m)^{-1}(\log n)^{-\frac{1}{2}}$ for a sufficiently large constant C, then the power of $\mathcal{R}_0 \cup \mathcal{R}_1$ is at least $1 - C_3 \log n/m^{1-2r}$.*

Proof. Note that for any $\boldsymbol{\theta} \in H_1'$, (136) implies that

$$\sum (|\theta_i| - c_m)_+ > e_m' = \kappa_m \gamma_m c_m m = Cm(\log n)^{-\frac{1}{2}}.$$

Then Lemma 5 leads to the desired result. ■

Lemma 9. *Consider testing H_0 (123) against H_1' (136) based on the rejection region $\mathcal{R}_0 \cup \mathcal{R}_2$. Suppose that $C_1(\log m/m)^{\frac{1}{4}} \leq c_m \leq (\log n)^{-\frac{1}{2}}$ and that $C_2 \log m \geq \log n$. If in (136), $\tilde{c}_m = (\gamma_m + 1)c_m$ for some $\gamma_m > 0$, $\kappa_m = C\gamma_m^{-1}$ for a sufficiently large constant C, then the power of $\mathcal{R}_0 \cup \mathcal{R}_2$ is at least $1 - C_3 n^{-\frac{1}{2}}$.*

Proof. Note that for any $\boldsymbol{\theta} \in H_1'$, (136) implies that

$$\sum_i (|\theta_i| - c_m)_+ > e_m' = \kappa_m \gamma_m c_m m = C m c_m.$$

Then Lemma 6 leads to the desired result. ■

Lemma 10. *Consider testing H_0 (123) against H_1' (136) based on the rejection region $\mathcal{R}_0 \cup \mathcal{R}_2$. Suppose that $0 < c_m \leq C_1(\log m/m)^{\frac{1}{4}}$ and that $C_2 \log m \geq \log n$. If in (136), $\tilde{c}_m = (\gamma_m + 1)C_1(\log m/m)^{\frac{1}{4}}$ for some $\gamma_m > 0$, $\kappa_m = C\gamma_m^{-1}$ for a sufficiently large constant C, then the power of $\mathcal{R}_0 \cup \mathcal{R}_2$ is at least $1 - C_3 n^{-\frac{1}{2}}$.*

Proof. Note that for any $\boldsymbol{\theta} \in H_1'$, (136) implies that

$$\sum_i (|\theta_i| - c_m)_+ > e_m' = \kappa_m \gamma_m c_m m = C C_1 m^{\frac{3}{4}} (\log m)^{\frac{1}{4}}.$$

Then Lemma 7 leads to the desired result. ■

8.2 Proof of Propositions 2 and 3

Proof of Proposition 2. After proper scaling by the factor $\sigma_n = \sigma n^{-\frac{1}{2}}$, Lemma 3 implies that for any j and any $j \leq l < J$, if $f \in H_{0,jl}$, then $P_f(R_{0,jl} \cup R_{1,jl}) \leq C' n^{-\frac{1}{2}}$. In addition, Lemma 4 further implies that for all $c_{jl} \leq \sigma_n (\log n)^{-\frac{1}{2}}$, when $f \in H_{0,jl}$, $P_f(R_{0,jl} \cup R_{2,jl}) \leq C' n^{-\frac{1}{2}}$. Combining the two bounds with the definition of ϕ_{jl} in (35), we obtain (56).

Turning to (57). Suppose for some (j, l), $f \in H_{1,jl}$. Let j satisfy (26) for some $\beta \in [\beta_0, 2\beta_0]$ and $M \in [1, M_0]$. We divide the discussion into three different cases as in (55).

In the first case, after proper scaling, we satisfy the condition of Lemma 5, which implies

$$P_f(R_{0,jl} \cup R_{1,jl}) \geq 1 - C' \log n \, 2^{-l(1-2r)} \geq 1 - C' 2^{-l/2}.$$

Here, the last inequality holds for all $r < 1/4$.

In the second case, after proper scaling, we satisfy the condition of Lemma 6. Thus, $P_f(R_{0,jl} \cup R_{2,jl}) \geq 1 - C' 2^{-l/2}$.

In the third case, after proper scaling, we satisfy the condition of Lemma 7. Thus, we also have $P_f(R_{0,jl} \cup R_{2,jl}) \geq 1 - C' 2^{-l/2}$. This completes the proof. ■

Proof of Proposition 3. Note that (61) is the same as (56), which has been proved in the proof of Proposition 2. In what follows, we focus on (62). Suppose for some (j, l), $f \in H_{1,jl}$. Let j satisfy (26) for some $\beta \in [\beta_0, 2\beta_0]$ and $M \in [1, M_0]$. We divide the discussion into three different cases as in (60).

In the first case, after proper scaling, we apply Lemma 8 with $\gamma_m = (\log n)^{-1/4}$ to obtain

$$P_f(R_{0,jl} \cup R_{1,jl}) \geq 1 - C' \log n \, 2^{-l(1-2r)} \geq 1 - C' 2^{-l/2},$$

where the last inequality holds for $r < 1/4$.

In the second case, after proper scaling, we apply Lemma 9 with $\gamma_m = 2^{\frac{1}{2}\beta_0(l-j)}$ to obtain $P_f(R_{0,jl} \cup R_{2,jl}) \geq 1 - C'2^{-l/2}$.

In the third case, after proper scaling, we satisfy the condition of Lemma 10. So, we apply Lemma 10 with $\gamma_m = (\log n)^{\frac{\beta_0}{2(4\beta_0+1)}} 2^{\frac{1}{8}(j^t - l)}$ to obtain $P_f(R_{0,jl} \cup R_{2,jl}) \geq 1 - C'2^{-l/2}$. This completes the proof. ■

9 Proofs of Other Results

9.1 Proof of Lemma 1

Proof. For a vector $\xi \in \mathbb{R}^d$, it is easy to check that

$$\int_{\mathbb{R}^d} \frac{\psi_\xi^2(y)}{\phi_0(y)} dy = \frac{1}{2}(e^{-\|\xi\|_2^2/\sigma^2} + e^{\|\xi\|_2^2/\sigma^2}) \leq \exp\left(\frac{1}{2\sigma^4}\|\xi\|_2^4\right).$$

Hence,

$$\int \frac{h_1(y)^2}{h_0(y)} = \int \prod_{i=1}^m \frac{\psi_{\gamma_{J_i}}^2(y_{J_i})}{\phi_0(y_{J_i})} = \prod_{i=1}^m \int \frac{\psi_{\gamma_{J_i}}^2(y_{J_i})}{\phi_0(y_{J_i})} \leq \prod_{i=1}^m \exp\left(\frac{1}{2\sigma^4}\|\gamma_{J_i}\|_2^4\right) = \exp\left(\frac{1}{2\sigma^4}\sum_{i=1}^m \|\gamma_{J_i}\|_2^4\right).$$

It then follows that

$$\chi^2(P_0, P_1) = \int \frac{h_1(y)^2}{h_0(y)} - 1 \leq \exp\left(\frac{1}{2\sigma^4}\sum_{i=1}^m \|\gamma_{J_i}\|_2^4\right) - 1.$$

Now note that if $\sum_{i=1}^m \|\gamma_{J_i}\|_2^4 \leq 2\sigma^4 \log(1 + \epsilon_0^2)$, then

$$\chi^2(P_0, P_1) \leq \exp\left(\frac{1}{2\sigma^4}\sum_{i=1}^m \|\gamma_{J_i}\|_2^4\right) - 1 \leq \epsilon_0^2.$$

Consequently, the L_1 distance between P_0 and P_1 satisfies

$$L_1(P_0, P_1) = \int |h_0 - h_1| \leq \chi(P_0, P_1) \leq \epsilon_0.$$

Hence if $P_0(A) \geq \alpha$, then $P_1(A) \geq P_0(A) - \frac{1}{2}L_1(P_0, P_1) \geq \alpha - \frac{1}{2}\epsilon_0$ and the lemma follows. ■

9.2 Proof of (14)

We first introduce a lemma regarding the difference between f and f_n. The proof is straightforward and can be found in, for instance, Cai (1996).

Lemma 11. *Suppose $f \in \Lambda(\beta, M)$ and the father wavelet ϕ has $s \geq \beta$ vanishing moments. Let $f_n(t) = \sum_{k=1}^n n^{-1/2} f(\frac{k}{n})\phi_{Jk}(t)$. Then there exist a constant c_ϕ depending only on the wavelet, such that*

$$\sup_t |f(t) - f_n(t)| \leq c_\phi M n^{-\beta}.$$

Proof of (14) Note that for any $l < J$ and any $1 \leq k \leq 2^l$,

$$
\begin{aligned}
|\bar{\theta}_{lk} - \theta_{lk}| &= |\int (f(t) - f_n(t))\psi_{lk}(t)dt| \\
&\leq \int |f(t) - f_n(t)||\psi_{lk}(t)|dt \leq c_\phi M n^{-\beta} \int 2^{l/2}|\psi(2^l - t)|dt \\
&\leq c_\phi M n^{-\beta} 2^{-l/2}\|\psi\|_1 \leq c_\psi M 2^{-(\beta+\frac{1}{2})l}.
\end{aligned}
$$

Here, the last inequality holds as $2^l \leq n$ for all $l < J$. This, together with (11), completes the proof. ■

References

[1] Cai, T. T. (1996). *Nonparametric function estimation via wavelets.* PhD Thesis, Cornell University.

[2] Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation.* Springer-Verlag, New York.