

# NONPARAMETRIC CONFIDENCE SETS FOR DENSITIES

Running Head: Nonparametric Confidence Sets

Woncheol Jang, Christopher Genovese, and Larry Wasserman  
Duke University and Carnegie Mellon University

We present a method for constructing nonparametric confidence sets for density functions based on an approach due to Beran and Dömbgen (1998). We expand the density in an appropriate basis and we estimate the basis coefficients by using linear shrinkage methods. We then find the limiting distribution of an asymptotic pivot based on the quadratic loss function. Inverting this pivot yields a confidence ball for the density.

AMS 2000 subject classification. Primary 62G07, 62G15; secondary 62G20.

Keywords and Phrases: Confidence Sets, nonparametric density estimation, shrinkage methods, empirical processes.

## 1 Introduction

This paper extends the REACT regression method of Beran (2000) and Beran and Dömbgen (1998) to the problem of density estimation. The goal is to obtain nonparametric confidence sets for density functions.

We expand the unknown density  $f$  in a basis,  $f(x) = \sum_j \theta_j \phi_j(x)$ , and we use an estimate of the form  $\hat{f}(x) = \sum_j \hat{\theta}_j \phi_j(x)$  where  $\hat{\theta}_j$  is an estimate of  $\theta_j$ . Based on the limiting distribution of  $\sum_j (\hat{\theta}_j - \theta_j)^2$  we construct a ball  $C_n$  such that

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}} P(f_p \in C_n) \geq 1 - \alpha \quad (1)$$

where  $f_p = \sum_{j=1}^p \theta_j \phi_j(x)$  is the projection of  $f$  onto the set spanned by the first  $p = p(n)$  basis functions and  $\mathcal{F}$  is an appropriate function class. Here,  $p = p(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . In the regression case, Beran and Dömbgen (1998) and Genovese and Wasserman (2002) used  $p(n) = n$ . However, in the density estimation case it appears we need the stronger condition  $p = o(n^{1/3})$ .

Constructing confidence sets for nonparametric curve estimation problems is challenging because the estimate  $\hat{f}$  typically has asymptotic bias. The

set  $C_n$  that we construct adjusts for some, but not all, of the bias. Specifically, the bias  $E(\widehat{f}(x)) - f(x)$  can be decomposed as  $(E(\widehat{f}(x)) - f_p(x)) + (f_p(x) - f(x))$  which we call smoothing bias and tail bias. The sets  $C_n$  account for smoothing bias but do not account for tail bias.

## 2 Review of the REACT Method

We begin with a brief review of Beran's REACT method for regression. Suppose that

$$Y_i = f(x_i) + \sigma \epsilon_i \quad (2)$$

where  $\epsilon_i \sim N(0, 1)$  and  $x_i = i/n$ . Expand  $f$  in an orthonormal basis as

$$f(x) = \sum_{j=1}^{\infty} \theta_j \phi_j(x). \quad (3)$$

Let  $\widehat{\theta}_j = 0$  for  $j > n$ , and for  $j \leq n$  define

$$\widehat{\theta}_j \equiv \frac{1}{n} \sum_{i=1}^n Y_i \phi_j(x_i). \quad (4)$$

Let  $\theta = (\theta_1, \dots, \theta_n)$  and let  $\widehat{\theta}(\lambda) = (\lambda_1 \widehat{\theta}_1, \dots, \lambda_n \widehat{\theta}_n)$  where the shrinkage coefficients  $\lambda = (\lambda_1, \dots, \lambda_n)$  are contained in an appropriate set  $\Lambda_n$  such as the set of all monotone, non-increasing vectors. Beran calls this the set of monotone modulators.

Define the *pivot process*

$$B_n(\lambda) \equiv \sqrt{n}(L_n(\lambda) - S_n(\lambda)) \quad (5)$$

where  $L_n(\lambda) = \sum_{j=1}^n (\widehat{\theta}_j - \theta_j)^2$  is the loss function and  $S_n(\lambda)$  is an estimate of the risk  $R_n(\lambda) = \mathbf{E}(L_n(\lambda))$  (such as Stein's unbiased risk estimator). The function estimate is

$$\widehat{f}(x) \equiv \sum_{j=1}^n \widehat{\lambda}_j \widehat{\theta}_j \phi_j(x)$$

where  $\widehat{\lambda}$  is the minimizer of  $S_n(\lambda)$  over  $\Lambda_n$ . Beran and Dümbgen (1998) showed that  $\{B_n(\lambda) : \lambda \in \Lambda_n\}$  converges to a Gaussian process. Moreover,

they showed that  $B_n(\widehat{\lambda})$  is stochastically close to  $B_n(\widetilde{\lambda})$  where  $\widetilde{\lambda}$  minimizes the true risk  $R_n(\lambda)$ . It follows that  $B_n(\widehat{\lambda})$  converges in law to a Normal with some variance  $\tau^2$ . The convergence is uniform over certain function classes  $\mathcal{F}$ . Then by inverting the pivot  $B_n(\widehat{\lambda})$  we get the confidence set

$$\mathcal{D}_n = \left\{ \theta : \sum_{j=1}^n (\theta_j - \widehat{\lambda}_j \widehat{\theta}_j)^2 \leq \frac{\widehat{\tau} z_\alpha}{\sqrt{n}} + S_n(\widehat{\lambda}) \right\}$$

where  $\widehat{\tau}$  is a consistent estimate of  $\tau$ . The corresponding confidence ball for  $f_n = \sum_{j=1}^n \theta_j \phi_j(x)$  is

$$\mathcal{C}_n = \left\{ f(\cdot) = \sum_{j=1}^n \theta_j \phi_j(\cdot) : (\theta_1, \dots, \theta_n) \in \mathcal{D}_n \right\}.$$

The theory described above was generalized to the case of non-linear wavelet thresholding where  $\mathcal{F}$  is a class of Besov spaces in Genovese and Wasserman (2002). However, the theory does not carry over directly to density estimation because the pivot process in density estimation is a sum of *dependent* processes. However, we will modify the theory to accommodate this complication.

### 3 Density Estimation

Let  $Y_1, \dots, Y_n$  be a random sample from a distribution function  $F$  with density  $f$  on  $[0, 1]$ . We assume that  $f \in L_2[0, 1]$  and we expand  $f$  in an orthonormal basis:

$$f(y) = 1 + \sum_{j=1}^{\infty} \theta_j \phi_j(y).$$

Although it is not essential, we shall use the cosine basis  $\phi_j(x) = \sqrt{2} \cos(j\pi y)$  for  $j = 1, 2, \dots$ . We estimate the density function by

$$\widehat{f}(y) \equiv 1 + \sum_{j=1}^p \widehat{\lambda}_j \widehat{\theta}_j \phi_j(y), \tag{6}$$

where  $\widehat{\theta}_j = \frac{1}{n} \sum_{i=1}^n \phi_j(Y_i)$  and  $p = p(n)$  depends on sample size  $n$ . Estimates of this form have been studied for a long time; see Efromovich (1999), for a detailed account.

We assume that  $f$  belongs to

$$\mathcal{F} \equiv \left\{ f : f(y) = 1 + \sum_{j=1}^{\infty} \theta_j \phi_j(y), \quad (\theta_1, \theta_2, \dots) \in \Theta(m, C) \right\},$$

where

$$\Theta(m, C) \equiv \left\{ (\theta_1, \theta_2, \dots) : \sum_{j=1}^{\infty} j^{2m} \theta_j^2 \leq C \right\}$$

is a Sobolev ellipsoid of order  $m$  and radius  $C$ . We assume that  $m > 1/2$ . However, we do not require that  $m$  or  $C$  be known. All results that follow hold uniformly over  $\Theta(m, C)$ .

Let

$$\Lambda_p = \left\{ (\lambda_1, \dots, \lambda_p) : 1 \geq \lambda_1 \geq \dots \geq \lambda_p \geq 0 \right\}$$

denote the set of monotone modulators. Let

$$L_p(\lambda) = \sum_{j=1}^p (\lambda_j \widehat{\theta}_j - \theta_j)^2, \tag{7}$$

be the loss function. The risk function is

$$R_p(\lambda) = \mathbb{E}(L_p(\lambda)) = \sum_{j=1}^p \left[ \lambda_j^2 \frac{\sigma_j^2}{n} + (1 - \lambda_j)^2 \theta_j^2 \right] \tag{8}$$

and an unbiased estimate of the risk is

$$S_p(\lambda) = \sum_{j=1}^p \left[ \lambda_j^2 \frac{\widehat{\sigma}_j^2}{n} + (1 - \lambda_j)^2 \left( \widehat{\theta}_j^2 - \frac{\widehat{\sigma}_j^2}{n} \right) \right], \tag{9}$$

where

$$\sigma_j^2 \equiv \text{Var}(\phi_j(Y_1)) = \left( 1 + \frac{\theta_{2j}}{\sqrt{2}} - \theta_j^2 \right), \quad \widehat{\sigma}_j^2 \equiv \frac{1}{n-1} \sum_{i=1}^n (\phi_j(Y_i) - \widehat{\theta}_j)^2.$$

The following theorem establishes the convergence of the pivot process

$$B_p(\lambda) = \sqrt{n}(L_p(\lambda) - S_p(\lambda)).$$

**THEOREM 3.1.** Suppose that  $\theta \in \Theta(m, C)$  where  $m > 1/2$ . Let  $p = p(n) \rightarrow \infty$  as  $n \rightarrow \infty$  and  $p = o(n^{1/3})$ . Further let  $\hat{\lambda}$  and  $\tilde{\lambda}$  be the minimizers of  $L_p(\lambda)$  and  $R_p(\lambda)$  over  $\Lambda_p$ . Then

$$\sqrt{n}(L_p(\hat{\lambda}) - S_p(\hat{\lambda}))/\hat{\tau}(\hat{\lambda}) \rightsquigarrow N(0, 1), \quad (10)$$

where  $\hat{\tau}^2(\hat{\lambda}) \equiv (p/n)\hat{\tau}_1^2(\hat{\lambda}) + \hat{\tau}_2^2(\hat{\lambda})$  is a consistent estimator of  $\tau^2(\tilde{\lambda}) \equiv (p/n)\tau_1^2(\tilde{\lambda}) + \tau_2^2(\tilde{\lambda})$ . Here,

$$\begin{aligned} \tau_1^2(\lambda) &\equiv \frac{2}{p} \sum_{j=1}^p (2\lambda_j - 1)^2 \sigma_j^4, & \hat{\tau}_1^2(\hat{\lambda}) &\equiv \frac{2}{p} \sum_{j=1}^p (2\hat{\lambda}_j - 1)^2 \hat{\sigma}_j^4, \\ \tau_2^2(\lambda) &\equiv 4 \sum_{j=1}^p \theta_j^2 (\lambda_j - 1)^2 \sigma_j^2 + 8 \sum_{1 \leq k < j \leq p} \theta_j \theta_k (\lambda_j - 1)(\lambda_k - 1) \sigma_{jk}, \\ \hat{\tau}_2^2(\hat{\lambda}) &\equiv 4 \sum_{j=1}^p \left( \hat{\theta}_j^2 - \frac{\hat{\sigma}_j^2}{n} \right) (\hat{\lambda}_j - 1)^2 \hat{\sigma}_j^2 + 8 \sum_{1 \leq k < j \leq p} \hat{\theta}_j \hat{\theta}_k (\hat{\lambda}_j - 1)(\hat{\lambda}_k - 1) \hat{\sigma}_{jk}, \\ \sigma_{jk} &\equiv \text{Cov}(\hat{\theta}_j, \hat{\theta}_k) = \left( \frac{\theta_{j+k} + \theta_{j-k}}{\sqrt{2}} - \theta_j \theta_k \right), & \hat{\sigma}_{jk} &\equiv \frac{1}{n-1} \sum_{i=1}^n (\phi_j(Y_i) - \hat{\theta}_j)(\phi_k(Y_i) - \hat{\theta}_k). \end{aligned}$$

The next theorem establishes how to construct uniform asymptotic confidence sets for densities. In what follows, we sometimes write  $P_f$  or  $P_\theta$  to emphasize the dependence of the probability measure on the unknown density. Define

$$\mathcal{D}_p = \left\{ \theta : \sum_{j=1}^p (\theta_j - \hat{\lambda}_j \hat{\theta}_j)^2 \leq \frac{z_\alpha \hat{\tau}(\hat{\lambda})}{\sqrt{n}} + S_p(\hat{\lambda}) \right\} \quad (11)$$

and

$$\mathcal{C}_p \equiv \left\{ f_p : f_p = 1 + \sum_{j=1}^p \theta_j \phi_j(x), \theta \in \mathcal{D}_p \right\}. \quad (12)$$

**THEOREM 3.2.** Under the same assumption of Theorem 3.1 the following results hold.

1.  $\hat{\tau}_k^2(\hat{\lambda})$  is a uniformly consistent estimator of  $\tau_k^2(\tilde{\lambda})$  for  $k = 1, 2$ :

$$\sup_{\theta \in \Theta(m, C)} P_\theta \left\{ \left| \hat{\tau}_k^2(\hat{\lambda}) - \tau_k^2(\tilde{\lambda}) \right| > \epsilon \right\} = o(1), \quad \forall \epsilon > 0. \quad (13)$$

2. The normalized pivot process  $B_p(\widehat{\lambda})/\tau(\widehat{\lambda})$  does not approach a degenerate distribution:

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta(m, C)} (\tau_1^2(\widetilde{\lambda}) + (n/p)\tau_2^2(\widetilde{\lambda})) > 0. \quad (14)$$

3. The confidence sets have coverage at least  $1 - \alpha$ :

$$\liminf_{n \rightarrow \infty} \inf_{\Theta(m, C)} P(\theta \in \mathcal{D}_p) \geq 1 - \alpha, \quad \text{and} \quad \liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}} P_f(f_p \in \mathcal{C}_p) \geq 1 - \alpha. \quad (15)$$

The proofs are in Section 4.

## 4 Proof of the Theorems

Throughout this section, all results hold uniformly over the Sobolev parameter ball  $\Theta(m, C)$  for  $m > 1/2$ . The following lemma plays a key role in the proof of Theorem 3.1.

LEMMA 4.1. *Suppose that  $f \in \mathcal{F}$  where  $m > 1/2$ . Then*

$$\sum_{j=0}^{\infty} |\theta_j|^k = O(1) \quad \text{for } k \geq 1.$$

Further let  $E_j^k = [\sqrt{n}(\widehat{\theta}_j - \theta_j)]^k$ . Then  $E_j^k = O_p(1)$  uniformly in  $j \in \{1, \dots, p\}$ .

PROOF. The first inequality is the Bernstein inequality for Fourier coefficients; see Efremovich (1999) for the details. For the bound of  $E_j^k$ , we can apply Serfling (1980) Section 2.2.2 Lemma B; the  $k$ th moment of the sum of IID random variables is of a order of  $O(n^{k/2})$  if  $k$ th moment of the random variable exists. Here  $n$  is the number of random variables. Since  $\phi_j(Y_1), \dots, \phi_j(Y_n)$  are IID and  $\mathbb{E}(\phi_j(Y_1) - \theta_j)^k$  is finite,

$$\mathbb{E} \left[ \sum_{j=1}^n (\phi_j(Y_1) - \theta_j) \right]^k = O(n^{k/2}), \quad \text{uniformly in } j.$$

□

Let  $E_j = \sqrt{n}(\widehat{\theta}_j - \theta_j)$  and define the standardized pivot process  $B_p(\lambda)/\tau(\lambda)$  as follows.

$$\begin{aligned} B_p(\lambda)/\tau(\lambda) &\equiv \sqrt{n}(L_p(\lambda) - S_p(\lambda))/\tau(\lambda) \\ &= [(p/n)^{1/2}W_1(\lambda) + W_2(\lambda) - (\sqrt{p}/n)V(\lambda)]/\tau(\lambda). \end{aligned}$$

Here,

$$\begin{aligned} W_1(\lambda) &\equiv \frac{1}{\sqrt{p}} \sum_{j=1}^p (2\lambda_j - 1)(E_j^2 - \sigma_j^2), \quad W_2(\lambda) \equiv 2 \sum_{j=1}^p (\lambda_j - 1)\theta_j E_j, \\ V(\lambda) &\equiv (n/p)^{1/2} \sum_{j=1}^p (2\lambda_j - 1)(\widehat{\sigma}_j^2 - \sigma_j^2), \quad \tau^2(\lambda) \equiv (p/n)\tau_1^2(\lambda) + \tau_2^2(\lambda), \end{aligned}$$

To prove Theorem 3.1, we follow the strategy of Beran and Dümbgen (1998), originating from Stein (1981). In other words, we use the asymptotic distribution of the pivot process  $B_p(\widehat{\lambda})/\tau(\widetilde{\lambda})$  to derive a confidence ball for  $\theta$ . Here,  $\widehat{\lambda}$  minimizes  $S_p(\lambda)$  and  $\widetilde{\lambda}$  minimizes  $R_p(\lambda)$ . The proof will use the following steps.

Step 1 Show that

$$(B_p(\widehat{\lambda}) - W(\widehat{\lambda}))/\tau(\widetilde{\lambda}) = o_p(1),$$

uniformly over  $\Theta(m, C)$ , where  $W(\lambda) \equiv (p/n)^{1/2}W_1(\lambda) + W_2(\lambda)$ .

Step 2 Show that  $W_k(\lambda)$  converges weakly to a Gaussian process with mean zero and covariance kernel  $K_k(s, t)$  where  $K_k(\lambda, \lambda) = \tau_k^2(\lambda)$  for  $k = 1, 2$ .

Step 3 Show that  $W(\widetilde{\lambda})/\tau(\widetilde{\lambda})$  converges to a standard Normal.

Step 4 Show that  $W(\widehat{\lambda})/\tau(\widetilde{\lambda})$  converges to a standard Normal by stochastic closeness of  $W_k(\lambda)$ .

$$\sup_{\Theta(m, C)} |W_k(\widehat{\lambda}) - W_k(\widetilde{\lambda})| = o_p(1) \quad \text{for } k = 1, 2.$$

Step 5 Show that  $B_p(\widehat{\lambda})/\widehat{\tau}^2(\widehat{\lambda})$  converges to a standard Normal where  $\widehat{\tau}^2(\widehat{\lambda})$  is a consistent estimator of  $\tau^2(\widetilde{\lambda})$ .

We begin by showing that

$$(B_p(\widehat{\lambda}) - W(\widehat{\lambda}))/\tau(\widetilde{\lambda}) = (\sqrt{p}/n)(V(\widehat{\lambda})/\tau(\widetilde{\lambda})) = o_p(1).$$

It suffices to show that

$$V(\widehat{\lambda}) = O(1) \quad \text{and} \quad \liminf_n \inf_{\theta \in \Theta(m,c)} n\tau(\widetilde{\lambda})/\sqrt{p} = \infty. \quad (16)$$

Recall that

$$\begin{aligned} \widehat{\sigma}_j^2 &\equiv \frac{1}{n-1} \sum_{i=1}^n (\phi_j(Y_i) - \widehat{\theta}_j)^2 \\ &= \frac{n}{n-1} \left( 1 + \frac{\widehat{\theta}_{2j}}{\sqrt{2}} - \widehat{\theta}_j^2 \right). \end{aligned}$$

Let  $V_j = \sqrt{n}(\widehat{\sigma}_j^2 - \sigma_j^2)$ .

Then,

$$\begin{aligned} V_j &= \sqrt{n} \left[ \frac{n}{n-1} \left( 1 + \frac{\widehat{\theta}_{2j}}{\sqrt{2}} - \widehat{\theta}_j^2 \right) - \left( 1 + \frac{\theta_{2j}}{\sqrt{2}} - \theta_j^2 \right) \right] \\ &= \frac{n}{n-1} \left[ \frac{\sqrt{n}(\widehat{\theta}_{2j} - \theta_{2j})}{\sqrt{2}} - \sqrt{n}(\widehat{\theta}_j^2 - \theta_j^2) \right] + \frac{\sqrt{n}}{n-1} \sigma_j^2 \\ &= \frac{n}{n-1} \left[ \frac{E_{2j}}{\sqrt{2}} - 2\theta_j E_j - \frac{E_j^2}{\sqrt{n}} \right] + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Thus,

$$\begin{aligned} \text{Var}(V(\lambda)) &= \frac{1}{p} \text{Var} \left( \sum_{j=1}^p (2\lambda_j - 1) V_j \right) \\ &= \frac{1}{p} \sum_{j=1}^p \sum_{k=1}^p (2\lambda_j - 1)^2 (2\lambda_k - 1)^2 \text{Cov}(V_j, V_k) \\ &\leq \frac{1}{p} \sum_{j=1}^p \sum_{k=1}^p |\text{Cov}(V_j, V_k)|. \end{aligned}$$

Here,  $\text{Cov}(V_j, V_k)$  is a linear combination of  $\text{Cov}(E_j, E_k)$ ,  $\text{Cov}(E_j, E_k^2)$  and  $\text{Cov}(E_j^2, E_k^2)$ .



To show  $\text{Var}(V(\lambda)) = O(1)$ , we need to show that sums of  $\text{Cov}(E_j, E_k)$ ,  $\text{Cov}(E_j, E_k^2)$  and  $\text{Cov}(E_j^2, E_k^2)$  are at most  $O(p)$ .

The followings are immediately from Lemma 4.1 :

$$\begin{aligned}
\sum_j^p \sum_k^p |\text{Cov}(E_j, E_k)| &= n \sum_j^p \sum_k^p |\text{Cov}(\widehat{\theta}_j, \widehat{\theta}_k)| \\
&= \sum_j^p \sum_k^p \left| \frac{\theta_{j+k} + \theta_{|j-k|}}{\sqrt{2}} - \theta_j \theta_k \right| \\
&= O(p), \\
\sum_j^p \sum_k^p |\text{Cov}(E_j^2, E_k)| &= n\sqrt{n} \sum_j^p \sum_k^p \left| \text{Cov}(\widehat{\theta}_j^2, \widehat{\theta}_k) - 2\theta_k \text{Cov}(\widehat{\theta}_j, \widehat{\theta}_k) \right| \\
&= \frac{1}{\sqrt{n}} \sum_j^p \sum_k^p \left| \frac{\theta_{2j+k} + \theta_{|2j-k|}}{2} - 2\theta_j \sigma_{jk} - \frac{\theta_k \theta_{2j}}{\sqrt{2}} \right| \\
&= O\left(\frac{p}{\sqrt{n}}\right), \\
\sum_j^p \sum_k^p |\text{Cov}(E_j^2, E_k^2)| &= n^2 \sum_j^p \sum_k^p |\text{Cov}(\widehat{\theta}_j^2 - 2\theta_j \widehat{\theta}_j, \widehat{\theta}_k^2 - 2\theta_k \widehat{\theta}_k)| \\
&= n^2 \sum_j^p \sum_k^p \left| \text{Cov}(\widehat{\theta}_j^2, \widehat{\theta}_k^2) - 2\theta_j \text{Cov}(\widehat{\theta}_j, \widehat{\theta}_k^2) \right. \\
&\quad \left. - 2\theta_k \text{Cov}(\widehat{\theta}_k, \widehat{\theta}_j^2) + 4\theta_k \theta_j \text{Cov}(\widehat{\theta}_j, \widehat{\theta}_k) \right| \\
&= \sum_j^p \sum_k^p \left| (\theta_{j+k} + \theta_{|j-k|} - \sqrt{2}\theta_j \theta_k)^2 + R_n \right| \\
&= 2 \sum_j^p \sum_k^p \sigma_{jk}^2 + O\left(\frac{p^2}{n}\right) = O(1),
\end{aligned}$$

where

$$R_n = \frac{1}{2\sqrt{2}n} \left( \theta_{2(j+k)} + \theta_{2|j-k|} + 2\theta_{2j} + 2\theta_{2k} + 2\sqrt{2} \right).$$

Consequently,  $\text{Var}(V(\lambda)) = O(1)$ .

We'll show in Lemma 4.4 that

$$\liminf_n \inf_{\Theta(m,C)} (\tau_1^2(\lambda) + (n/p)\tau_2^2(\lambda)) > 0.$$

Hence,

$$\liminf_n \inf_{\Theta(m,C)} n^2 \tau^2(\tilde{\lambda})/p = \liminf_n \inf_{\Theta(m,C)} n \tau_1^2(\tilde{\lambda}) + (n^2/p) \tau_2^2(\tilde{\lambda}) = \infty.$$

Therefore,

$$\limsup_n \sup_{\Theta(m,C)} (\sqrt{p}/n) (V(\hat{\lambda})/\tau(\tilde{\lambda})) = 0.$$

The next step is to show that  $W(\hat{\lambda})/\tau(\tilde{\lambda})$  converges to a standard Normal. Now,

$$W(\hat{\lambda})/\tau(\tilde{\lambda}) = (a(\tilde{\lambda})/\tau_1^2(\tilde{\lambda}))^{1/2} W_1(\hat{\lambda}) + ((1 - a(\tilde{\lambda}))/\tau_2^2(\tilde{\lambda}))^{1/2} W_2(\hat{\lambda}),$$

where  $a(\lambda) = (p/n) \tau_1^2(\lambda)/\tau^2(\lambda)$ . Note that  $a(\lambda) \in [0, 1]$ .

First, we derive the asymptotic distribution of  $W(\tilde{\lambda})/\tau(\tilde{\lambda})$  using the following strategy.

Step 1 Show that the characteristic functions(c.fs) of  $(a(\tilde{\lambda})/\tau_1^2(\tilde{\lambda}))^{1/2} W_1(\tilde{\lambda})$  and  $((1 - a(\tilde{\lambda}))/\tau_2^2(\tilde{\lambda}))^{1/2} W_2(\tilde{\lambda})$  converges to the c.fs of Normal distributions. That is,

$$\begin{aligned} \left| \mathbb{E} \left[ \exp \left( it \left( \frac{a(\tilde{\lambda})}{\tau_1^2(\tilde{\lambda})} \right)^{1/2} W_1(\tilde{\lambda}) \right) \right] - \exp \left( - \frac{t^2 a(\tilde{\lambda})}{2} \right) \right| &= o(1), \\ \left| \mathbb{E} \left[ \exp \left( it \left( \frac{(1 - a(\tilde{\lambda}))}{\tau_2^2(\tilde{\lambda})} \right)^{1/2} W_2(\tilde{\lambda}) \right) \right] - \exp \left( - \frac{t^2 (1 - a(\tilde{\lambda}))}{2} \right) \right| &= o(1), \end{aligned}$$

if

$$\liminf_n \inf_{\Theta(m,C)} \tau_1^2(\tilde{\lambda})/a(\tilde{\lambda}) > 0, \quad \liminf_n \inf_{\Theta(m,C)} \tau_2^2(\tilde{\lambda})/(1 - a(\tilde{\lambda})) > 0. \quad (17)$$

Step 2 Show asymptotic independence of  $(a(\tilde{\lambda})/\tau_1^2(\tilde{\lambda}))^{1/2} W_1(\tilde{\lambda})$  and  $((1 - a(\tilde{\lambda}))/\tau_2^2(\tilde{\lambda}))^{1/2} W_2(\tilde{\lambda})$ .

In other words,

$$\sup_{\Theta(m,C)} \left( \frac{a(\tilde{\lambda})(1 - a(\tilde{\lambda}))}{\tau_1^2(\tilde{\lambda})\tau_2^2(\tilde{\lambda})} \right)^{1/2} \text{Cov}(W_1(\tilde{\lambda}), W_2(\tilde{\lambda})) = o(1) \quad (18)$$

If conditions (17) and (18) are satisfied, then

$$\left| \mathbb{E} \left[ \exp \left( it \left( \frac{a(\tilde{\lambda})}{\tau_1^2(\tilde{\lambda})} \right)^{1/2} W_1(\tilde{\lambda}) + \left( \frac{1-a(\tilde{\lambda})}{\tau_2^2(\tilde{\lambda})} \right)^{1/2} W_2(\tilde{\lambda}) \right) \right] - \exp \left( -\frac{t^2}{2} \right) \right| = o(1).$$

Hence,  $W(\tilde{\lambda})/\tau(\tilde{\lambda})$  converges to a standard Normal.

Note that conditions (17) and (18) can be replaced with

$$\sup_{\Theta(m,C)} \text{Cov}(W_1(\tilde{\lambda}), W_2(\tilde{\lambda})) = o(1), \quad (19)$$

$$\liminf_n \inf_{\Theta(m,C)} \left( \frac{\tau_1^2(\tilde{\lambda})\tau_2^2(\tilde{\lambda})}{a(\tilde{\lambda})(1-a(\tilde{\lambda}))} \right)^{1/2} > 0. \quad (20)$$

To show the asymptotic normality of  $W(\hat{\lambda})/\tau(\hat{\lambda})$ , it suffices to show  $W(\hat{\lambda})/\tau(\hat{\lambda})$  is stochastically close to  $W(\tilde{\lambda})/\tau(\tilde{\lambda})$ . Since  $a(\lambda)$  is bounded, one can show the stochastic closeness by showing stochastic closeness of  $W_k(\tilde{\lambda})/\tau_k(\tilde{\lambda})$  to  $W_k(\hat{\lambda})/\tau_k(\hat{\lambda})$  for  $k = 1, 2$  :

$$\sup_{\Theta(m,C)} |W_k(\tilde{\lambda})/\tau_k(\tilde{\lambda}) - W_k(\hat{\lambda})/\tau_k(\hat{\lambda})| = o_p(1). \quad (21)$$

To show the stochastic closeness, we invoke Theorem 6.2 in the Appendix which is a modified functional Central limit theorem. To do so, we must show finite dimensional convergence of  $W_1$  and  $W_2$  to a Gaussian limit which also guarantees that the c.f.'s (characteristic functions) of  $W_1$  and  $W_2$  converges to the c.f.'s of Normal distributions.

We use the following lemma to show finite dimensional convergence of  $W_1$  to a Gaussian limit.

LEMMA 4.2. *For given  $\lambda$ ,*

$$\left| \mathbb{E} \left( \exp[itW_1(\lambda)] \right) - \exp \left( -\frac{t^2\tau_1^2(\lambda)}{2} \right) \right| \rightarrow 0, \quad n \rightarrow 0. \quad (22)$$

*Furthermore, if*

$$\liminf_n \inf_{\Theta(m,C)} \tau_1^2(\lambda) > 0, \quad (23)$$

*then the finite dimensional distribution of  $W_1(\lambda)$  has a Gaussian limit.*

PROOF.

First we show that the summability of covariance condition (31) is satisfied.

Let  $W_1(\lambda) = \sum_{j=1}^p S_j(\lambda)$  where  $S_j(\lambda) = \frac{1}{\sqrt{p}}(2\lambda_j - 1)W_{1j}$ .

Then,

$$\begin{aligned} \left| \sup_{s,t \in \Lambda} \sum_{1 \leq k < j \leq p} \mathbb{E}(S_j(s)S_k(t)) \right| &\leq \frac{1}{p} \sup_{s,t \in \Lambda} \sum_{1 \leq k < j \leq p} \left| (2s_j - 1)(2t_k - 1) \mathbb{E}[W_{1j}W_{1k}] \right| \\ &= \frac{1}{p} \sum_{1 \leq k < j \leq p} \left| \mathbb{E}[(E_j^2 - \sigma_j^2)(E_k^2 - \sigma_k^2)] \right| \\ &\leq \frac{1}{p} \sum_{1 \leq k < j \leq p} |\text{Cov}(E_j^2, E_k^2)|. \end{aligned}$$

We already showed that  $\text{Cov}(E_j^2, E_k^2) = \sigma_{jk}^2 + O(n^{-1})$ . Therefore,

$$\frac{1}{p} \sum_{1 \leq j \neq k \leq p} |\text{Cov}(E_j^2, E_k^2)| = o(1). \quad (24)$$

Now we show finite dimensional the convergence of the finite dimensional marginals of  $W_1(\lambda)$ . One can write  $W_1(\lambda)$  as a sum of three terms.

$$W_1(\lambda) = \frac{w_1}{\sqrt{p_1}} \sum_{j \in J_1} (2\lambda_j - 1)W_{1j} + \frac{w_2}{\sqrt{p_2}} \sum_{j \in J_2} (2\lambda_j - 1)W_{1j} - \frac{w_3}{\sqrt{p_3}} \sum_{j \in J_3} (1 - 2\lambda_j)W_{1j},$$

where  $J_1 = \{j : (2\lambda_j - 1) \geq 0\}$ ,  $J_2 = \{j : (2\lambda_j - 1) = 0\}$ ,  $J_3 = \{j : (2\lambda_j - 1) < 0\}$ ,  $w_k = (p_k/p)^{1/2}$ ,  $p_k$  is the cardinality of  $J_k$  for  $k = 1, 2, 3$  and  $\sum_k p_k = p$ .

Let  $W_{n,k} \equiv \frac{1}{\sqrt{p_k}} \sum_{j \in J_k} (2\lambda_j - 1)W_{kj}$ .

Then,

$$W_1 = w_1 W_{n,1} + w_3 W_{n,3},$$

since  $W_{n,2} = 0$ .

To show the asymptotic normality of  $W_1(\lambda)$ , we need to show (1) the c.fs of  $W_{n,1}$  and  $W_{n,3}$  converges to the c.fs of Normal distributions (2) conditions (19) and (20) are satisfied :

$$\sup_{\Theta(m,C)} \text{Cov}(W_{n,1}(\tilde{\lambda}), W_{n,3}(\tilde{\lambda})) = o(1), \quad \liminf_n \inf_{\Theta(m,C)} \frac{1}{w_1} \frac{1}{w_2} > 0, \quad (25)$$

(3)  $W_1(\lambda)$  does not approach a degenerated distribution.

Condition (23) ensures (3). Without (3), (1) and (2) imply only the convergence of the characteristic function of  $W_1(\lambda)$ . Keep in mind that we only need convergence of the c.f of  $W_1(\lambda)$  to the c.f. of a Normal distribution to invoke Theorem 6.2.

Since  $J_1$  and  $J_3$  are disjoint, one can show that

$$\sup_{\Theta(m,C)} \text{Cov}(W_{n,1}(\tilde{\lambda}), W_{n,3}(\tilde{\lambda})) = o(1) \quad (26)$$

and

$$\liminf_n \inf_{\Theta(m,C)} \frac{1}{w_1} \frac{1}{w_3} > 0, \quad (27)$$

because  $w_1, w_2 \in [0, 1]$ .

It remains to show convergence of the characteristic functions of  $W_{n,1}$  and  $W_{n,3}$ . Without loss of generality, it suffices to show that the c.f. of  $W'_1(a) \equiv \frac{1}{\sqrt{p}} \sum_{j=1}^p a_j W_{1j}$  converges to the c.f of a Normal distribution where  $a = (a_1, \dots, a_p)$  and  $a_j$ 's are bounded and positive. While Theorem 3.1 requires the finite dimensional convergence, one can reduce it to the univariate case via linear combinations by Cramér-Wold device. In other words, the c.f. of  $\sum_{k=1}^m W_1(t_k) = \frac{1}{\sqrt{p}} \sum_{j=1}^p \sum_{k=1}^m (2t_{kj} - 1) W_{1j}$  converges to the c.f. of a Normal distribution where  $t_k = (t_{k1}, \dots, t_{kp}) \in \Lambda_p$ .

Since  $\sum_{k=1}^m (2t_{kj} - 1)$  is bounded for all  $j$ , again it remains to show convergence of the c.f. of  $W'_1(a)$  to the characteristic function of a Normal distribution.

Let  $X_i = (\phi_1(Y_i), \dots, \phi_p(Y_i))^T$  and  $D = \text{diag}(a_1, \dots, a_p)$ . Further let  $R_i = \Sigma^{-1/2}(X_i - \theta)$  where  $\Sigma \equiv \text{Var}(X_i)$ . Then  $W'_1(a)$  can be written as follows:

$$\begin{aligned} W'_1(a) &= \frac{1}{\sqrt{p}} \sum_{j=1}^p a_j \left( [\sqrt{n}(\hat{\theta}_j - \theta_j)]^2 - \sigma_j^2 \right) \\ &= \frac{1}{\sqrt{p}} (\sqrt{n}(\bar{X} - \theta))^T D (\sqrt{n}(\bar{X} - \theta)) - \frac{1}{\sqrt{p}} \sum_{j=1}^p a_j \sigma_j^2 \\ &= \frac{1}{\sqrt{p}} (\sqrt{n}\bar{R})^T \Sigma^{1/2} D \Sigma^{1/2} (\sqrt{n}\bar{R}) - \frac{1}{\sqrt{p}} \sum_{j=1}^p a_j \sigma_j^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{p}}(\sqrt{n}\bar{R})^T \Sigma'(\sqrt{n}\bar{R}) - \frac{1}{\sqrt{p}} \sum_{j=1}^p a_j \sigma_j^2 \\
&= \frac{1}{\sqrt{p}}(\sqrt{n}\bar{R})^T D'(\sqrt{n}\bar{R}) - \frac{1}{\sqrt{p}} \sum_{j=1}^p a_j \sigma_j^2 + \frac{1}{\sqrt{p}}(\sqrt{n}\bar{R})^T (\Sigma' - D')(\sqrt{n}\bar{R}),
\end{aligned}$$

where  $\Sigma' \equiv \Sigma^{1/2} D \Sigma^{1/2}$  and  $D' \equiv \text{diag}(\Sigma')$ .

Here,

$$D' = \text{diag}(\Sigma^{1/2}) \cdot D \cdot \text{diag}(\Sigma^{1/2}) = \text{diag}(\Sigma) D = \text{diag}(a_1 \sigma_1^2, \dots, a_p \sigma_p^2).$$

With the covariance summability condition

$$\text{Var} \left( \frac{1}{\sqrt{p}} [(\sqrt{n}\bar{R})^T (\Sigma' - D')(\sqrt{n}\bar{R})] \right) = \frac{2}{p} \sum_{1 \leq j < k \leq p} a_j a_k \text{Cov}(E_j^2, E_k^2) = o(1).$$

As a result,

$$W'(a) = h(\sqrt{n}\bar{R}) + o_p(1),$$

where  $h(x) = \frac{1}{\sqrt{p}}(x^T D' x - 1^T D' 1)$  and  $x = (x_1, \dots, x_p)$ ,  $1 = (1, \dots, 1)^T \in \mathbb{R}^p$ .

We only need to show convergence of the moment generating function (mgf) of  $h(\sqrt{n}\bar{R})$  to the mgf of a Normal distribution, that is, we must show that

$$|\mathbb{E} e^{-sh(\sqrt{n}\bar{R})} - e^{s^2 \sum_{j=1}^p a_j^2 \sigma_j^4 / p}| = o(1), \quad \text{for any } s > 0.$$

By the triangle inequality,

$$|\mathbb{E} e^{-sh(\sqrt{n}\bar{R})} - e^{s^2 \sum_{j=1}^p a_j^2 \sigma_j^4 / p}| \leq |\mathbb{E} (e^{-sh(\sqrt{n}\bar{R})}) - \mathbb{E} (e^{-sh(Z)})| + |\mathbb{E} (e^{-sh(Z)}) - e^{s^2 \sum_{j=1}^p a_j^2 \sigma_j^4 / p}|,$$

where  $Z = (Z_1, \dots, Z_p)^T$  is a multivariate standard Normal random variable.

In addition to the covariance summability condition, convergence of the first term of the right hand side implies asymptotic independence while that of the second term ensures that the sum of independent random variables converges to a Normal distribution.

To show convergence of the first term, we adapt Portnoy's (1986) approach and use the Lindeberg condition for the proof of the convergence of

the second term. We'll show in Lemma 4.3 that the Lindeberg condition is satisfied.

The Fourier inversion formula gives

$$\begin{aligned} \left| \mathbb{E}(e^{-sh(\sqrt{n}\bar{R})}) - \mathbb{E}(e^{-sh(Z)}) \right| &= \left| \int g(x) dP_n(x) - \int g(x) d\Phi(x) \right| \\ &= \left| (2\pi)^{-p} \int \hat{g}(t) \left( \varphi_R^n\left(\frac{t}{\sqrt{n}}\right) - e^{-\frac{\|t\|^2}{2}} \right) dt \right|, \end{aligned}$$

where  $g(x) = e^{-sh(x)}$ ,  $\hat{g} = \int e^{-it^T x} g(x) dx$ ,  $P_n$  is the cumulative density function (cdf) of  $\sqrt{n}\bar{R}$  and  $\Phi(x)$  is the cdf of  $Z$ .

The key point is to use properties of  $\hat{g}(t)$  to show that the above integral converges to 0.

Let  $\beta_j^2 = \frac{\sqrt{p}}{2sa_j\sigma_j^2}$ . Then,

$$\begin{aligned} \hat{g}(t) &= \int e^{-it^T x} g(x) dx \\ &= \exp\left(\frac{s}{\sqrt{p}} 1^T D' 1\right) \int \exp(-it^T x) \exp\left(-\frac{s}{\sqrt{p}} x^T D' x\right) dx \\ &= (2\pi)^{p/2} \left(\prod_{j=1}^p \beta_j^2\right)^{1/2} \exp\left(\frac{1}{2} \sum_{j=1}^p \frac{1}{\beta_j^2} - \frac{1}{2} \sum_{j=1}^p t_j^2 \beta_j^2\right). \end{aligned}$$

Note that  $\hat{g}(t)$  converges to 0 exponentially fast unless  $\|t\|$  is very small.

Now,

$$\varphi_R\left(\frac{t}{\sqrt{n}}\right) = \mathbb{E}\left(\exp\left(\frac{t^T}{\sqrt{n}} R\right)\right) = \mathbb{E}\left(\exp\left(\frac{\|t\|}{\sqrt{n}} \frac{t^T}{\|t\|} R\right)\right) = \varphi_{R'}\left(\frac{\|t\|}{\sqrt{n}}\right),$$

where  $R' = \frac{t^T}{\|t\|} R$ . Note  $\text{Var}(R') = 1$ .

By a Taylor expansion,

$$\varphi_R\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{\|t\|^2}{2n} - \frac{i\mathbb{E}(t^T R)^3}{6n\sqrt{n}} + e(t), \quad |e(t)| \leq \sup_t \mathbb{E}(t^T R)^4 / (24n^2).$$

Let  $s^T = t^T \Sigma^{-1/2}$ . Then  $s^T \Sigma s = \|t\|^2$ .

Furthermore,

$$\begin{aligned}
\mathbb{E} (t^T R)^4 &= \mathbb{E} [s^T (X - \theta)]^4 \\
&= \mathbb{E} \left[ \sum_{j=1}^p s_j (\varphi_j(Y) - \theta_j) \right]^4 \\
&\leq 8 \left( \sum_{j=1}^p s_j \right)^2 \mathbb{E} \left[ \sum_{j=1}^p s_j (\varphi_j(Y) - \theta_j) \right]^2 \\
&= 8 \left( \sum_{j=1}^p s_j \right)^2 s^T \Sigma s.
\end{aligned}$$

because  $|\varphi_j(Y) - \theta_j| \leq 2\sqrt{2}$ .

Hence,

$$\frac{1}{\|t\|^4} \mathbb{E} (t^T R)^4 \leq 8 \left( \sum_{j=1}^p s_j \right)^2 s^T \Sigma s / (s^T \Sigma s)^2 = 8 \left( \sum_{j=1}^p s_j \right)^2 / s^T \Sigma s = O(\|t\|^2).$$

Define  $\Psi_n(t) = \frac{\|t\|^2}{2} + i \mathbb{E} (t^T R)^3 / (6\sqrt{n})$ . Then,

$$\left| \log \varphi_R^n \left( \frac{t}{\sqrt{n}} \right) - \Psi_n(t) \right| \leq C \frac{\|t\|^6}{n}.$$

As a result, one can obtain the following from the fact  $|e^u - 1| \leq |u|e^{|u|}$ ,

$$\left| \varphi_R^n \left( \frac{t}{\sqrt{n}} \right) - e^{-\Psi_n(t)} \right| \leq C \frac{\|t\|^6}{n} \exp \left( - \frac{\|t\|^2}{2} \left( 1 - \frac{2C\|t\|^4}{n} \right) \right).$$

Again applying the triangle inequality gives

$$\begin{aligned}
\left| (2\pi)^{-p} \int \widehat{g}(x) \left( \varphi_R^n \left( \frac{t}{\sqrt{n}} \right) - e^{-\frac{\|t\|^2}{2}} \right) dt \right| &\leq (2\pi)^{-p} \int |\widehat{g}(x)| \left| \varphi_R^n \left( \frac{t}{\sqrt{n}} \right) - e^{-\Psi_n(t)} \right| dt \\
&\quad + (2\pi)^{-p} \int |\widehat{g}(x)| \left| 1 - e^{-i \mathbb{E} (t^T R)^3 / (6\sqrt{n})} \right| e^{-\frac{1}{2} \|t\|^2} dt.
\end{aligned}$$

The second term on right hand side converges to 0 from Portnoy (1988) Lemma 1.1. For the first term, one can split the domain of the integral into  $A_1 = \{\|t\| \leq \epsilon p^{1/3}\}$  and  $A_2 = \{\|t\| > \epsilon p^{1/3}\}$  and define  $I_1$  and  $I_2$  are the integrals over  $A_1$  and  $A_2$ .



Now,

$$\begin{aligned} I_1 &\leq (2\pi)^{-p} \int |\widehat{g}(t)| \frac{Cp^3\epsilon^3}{n} \exp\left(-\frac{\|t\|^2}{2} + \frac{Cp^3\epsilon^3}{n}\right) \\ &\leq (2\pi)^{-p/2} \mathbb{E}(\widehat{g}(Z)) \frac{Cp^3\epsilon^3}{n} \exp\left(\frac{Cp^3\epsilon^3}{n}\right). \end{aligned}$$

where  $Z$  is a multivariate standard Normal random variable.

Furthermore,

$$\begin{aligned} (2\pi)^{-p/2} \mathbb{E}(\widehat{g}(Z)) &= (2\pi)^{-p} \left(\prod_{j=1}^p \beta_j^2\right)^{1/2} \exp\left(\frac{1}{2} \sum_{j=1}^p \frac{1}{\beta_j^2}\right) \int \exp\left(\frac{1}{2} \sum_{j=1}^p t_j^2 (1 + \beta_j^2)\right) dt \\ &= (2\pi)^{p/2} \exp\left(\frac{1}{2} \sum_{j=1}^p \frac{1}{\beta_j^2}\right) \left(\prod_{j=1}^p \left(\frac{\beta_j^2}{1 + \beta_j^2}\right)\right)^{1/2} + o(1) \\ &= (2\pi)^{p/2} \prod_{j=1}^p \left(1 + \frac{1}{2\beta_j^2} + \frac{1}{8\beta_j^4}\right) \prod_{j=1}^p \left(1 - \frac{1}{2\beta_j^2} + \frac{3}{8\beta_j^4}\right) + o(1) \\ &= (2\pi)^{p/2} \prod_{j=1}^p \left(1 + \frac{1}{4\beta_j^4}\right) + o(1) \\ &= (2\pi)^{-p/2} \exp\left(\frac{1}{4} \sum_{j=1}^p \frac{1}{\beta_j^4}\right) + o(1) \\ &= (2\pi)^{-p/2} \exp\left(\frac{1}{p} \sum_{j=1}^p s^2 a_j^2 \sigma_j^4\right) + o(1). \end{aligned}$$

which is the mgf of a Normal distribution.

Consequently, with  $p^3/n \rightarrow 0$ ,  $I_1$  converges to 0.

For  $t \in A_2$ ,

$$\begin{aligned} \widehat{g}(t) &= (2\pi)^{p/2} \left(\prod_{j=1}^p \beta_j^2\right)^{1/2} \exp\left(\frac{1}{2} \sum_{j=1}^p \frac{1}{\beta_j^2} - \frac{1}{2} \sum_{j=1}^p t_j^2 \beta_j^2\right) \\ &\geq (2\pi)^{p/2} \exp(C_1 p \log p + C_2 \sqrt{p} - C_3 p^{7/6}). \end{aligned}$$

because  $\sum_{j=1}^p \beta_j^2 t_j^2 \leq \max_j \beta_j^2 \|t\|^2 = O(p^{7/6})$ .

Then,

$$\begin{aligned}
I_2 &\leq \sup_{t \in A_2} |\widehat{g}(t)| (2\pi)^{-p/2} \int \left| \varphi_R^n \left( \frac{t}{\sqrt{n}} \right) \right| dt + \sup_{t \in A_2} |\widehat{g}(t)| (2\pi)^{-p/2} \int e^{-\Psi(t)} dt \\
&\leq \sup_{t \in A_2} |\widehat{g}(t)| (2\pi)^{-p/2} n^{p/2} \int |\varphi_R^n(t)| dt + \sup_{t \in A_2} |\widehat{g}(t)| (2\pi)^{-p/2} \int e^{-\frac{1}{2}\|t\|^2} dt.
\end{aligned}$$

Since  $\sup_{t \in A_2} |\widehat{g}(t)|$  converges to 0 and  $e^{-\|t\|^2/2}$  is integrable, the second term converges to 0.

For the first term, if  $\varphi_R^n(t)$  is integrable, then it is bounded by

$$C_0(2\pi)^{p/2} \exp(C_1 p \log p + C_2 \sqrt{p} - C_3 p^{7/6} + (p/2) \log n),$$

which converges to 0 due to  $p^3/n \rightarrow 0$ .

It remains to be shown that  $\varphi_R^n(t)$  is integrable. Since  $|\varphi_R(t)| \leq 1$ , it suffices to show that  $\varphi_R(t)$  is integrable.

Because the parameter of the probability density function (pdf) of  $Y$  belongs to the Sobolev parameter ball, the pdf is bounded. Recall  $X = (\phi_1(Y), \dots, \phi_p(Y))$ . Hence, the pdf of  $X$  is also bounded. Furthermore,  $R = \Sigma^{-1/2}(X - \theta)$  is a orthonormal transformation of  $X$ , therefore its pdf is also bounded. Since the pdf of  $R$  is bounded,  $\varphi_R(t)$  is integrable.

Finally, it remains to show  $\text{Var}(W_1(\lambda)) = \tau_1^2(\lambda) + o_p(1)$  for given  $\lambda$  :

$$\begin{aligned}
\text{Var}(W_1(\lambda)) &= \frac{1}{p} \text{Var} \left( \sum_{j=1}^p (2\lambda_j - 1) W_{1j} \right) \\
&= \frac{1}{p} \sum_{j=1}^p (2\lambda_j - 1)^2 \text{Var}(E_j^2) + \frac{2}{p} \sum_{1 \leq k < j \leq p} (2\lambda_j - 1)(2\lambda_k - 1) \text{Cov}(E_j^2, E_k^2) \\
&= \frac{2}{p} \sum_{j=1}^p (2\lambda_j - 1)^2 \sigma_j^4 + O\left(\frac{1}{p}\right), \\
&= \tau_1^2(\lambda) + o(1).
\end{aligned}$$

□

LEMMA 4.3.  $W_1(\lambda)$  and  $W_2(\lambda)$  converge weakly to Gaussian processes over  $\Lambda_p$ .

PROOF. Having proved finite dimensional convergence to a Gaussian limit and covariance summability of  $W_1(\lambda)$  in Lemma 4.2, we can now appeal to Theorem 6.2.

By direct calculation,

$$\|S_j(\lambda)\|_{\Lambda_p} = \frac{1}{\sqrt{p}} \sup_{\Lambda_p} |(2\lambda_j - 1)W_{1j}| \leq \frac{1}{\sqrt{p}} |W_{1j}|.$$

Applying Lemma 4.1 immediately implies  $W_{1j}^k = (E_j^2 - \sigma_j^2)^k = O_p(1)$ . Then,

$$\sum_{j=1}^p \mathbb{E} \left( \|S_j(\lambda)\|_{\Lambda_p}^2 \right) \leq \frac{1}{p} \sum_{j=1}^p \mathbb{E} (W_{1j})^2 = O(1).$$

Now, for all  $u > 0$ ,

$$\begin{aligned} \mathbb{E} \left( \sum_{j=1}^p I\{\|S_j(\lambda)\|_{\Lambda_p}^2 > u\} \|S_j(\lambda)\|_{\Lambda_p}^2 \right) &\leq \frac{1}{p} \sum_{j=1}^p \mathbb{E} \left( I\{W_{1j}^2 > up\} W_{1j}^2 \right) \\ &\leq \frac{1}{p} \sum_{j=1}^p (\mathbb{E} I\{W_{1j}^2 > up\} \mathbb{E} [W_{1j}^4])^{1/2} \\ &\leq \frac{1}{p^{3/2} \sqrt{u}} \sum_{j=1}^p (E[W_{1j}^2] E[W_{1j}^4])^{1/2} \\ &= O\left(\frac{1}{\sqrt{p}}\right). \end{aligned}$$

It remains to show that the entropy condition (36) is satisfied.

Let  $d_P(s, t) = (\int (s - t)^2 dP)^{1/2}$  where  $P$  is a random measure and

$$dP(\cdot) \equiv \begin{cases} \frac{\sum_{j=1}^p \delta_{t_j}(\cdot) W_{1j}^2}{\sum_{j=1}^p W_{1j}^2} & : \sum_{j=1}^p W_{1j}^2 > 0 \quad \text{a.e.} \\ 0 & : \text{otherwise.} \end{cases}$$

Here,

$$\delta_{t_j}(s) \equiv \begin{cases} 1 & : s = t_j \\ 0 & : \text{otherwise.} \end{cases}$$

Then, for arbitrary  $s, t \in \Lambda_p$ ,

$$\widehat{\rho}_{W_1}(s, t)^2 = \frac{4}{p} \sum_{j=1}^p \left[ (s_j - t_j) W_{1j} \right]^2$$

$$\leq \left[ d_P(t, s) \right]^2 \frac{4}{p} \sum_{j=1}^p W_{1j}^2.$$

Using the above inequality, one can make the covering number bounded by the uniform covering number:

$$N(u, \Lambda_p, \widehat{\rho}_{W_1}) \leq N \left( u \left[ \frac{4}{p} \sum_{j=1}^p W_{1j}^2 \right]^{-1/2}, \Lambda_p \right).$$

It follows that

$$\begin{aligned} \int_0^{\epsilon(n)} \sqrt{\log N(u, \Lambda_p, \widehat{\rho}_{W_1})} du &\leq \int_0^{\epsilon(n)} \left\{ \log N \left( u \left[ \frac{4}{p} \sum_{j=1}^p W_{1j}^2 \right]^{-1/2}, \Lambda_p \right) \right\}^{1/2} du \\ &= \left[ \frac{4}{p} \sum_{j=1}^p W_{1j}^2 \right]^{1/2} \int_0^{\epsilon(n) \left[ (4/p) \sum_{j=1}^p W_{1j}^2 \right]^{1/2}} \sqrt{\log N(u, \Lambda_p)} du. \end{aligned}$$

It is shown by Dudley (1987) that

$$\log N(u, \Lambda_p) \leq cu^{-1} \quad \text{for all } u \in (0, 1].$$

Combining the above inequality and  $\frac{4}{p} \sum_{j=1}^p W_{1j}^2 = O_p(1)$  gives

$$\int_0^{\epsilon(n) \left[ (4/p) \sum_{j=1}^p W_{1j}^2 \right]^{1/2}} \sqrt{\log N(u, \Lambda_p)} du \rightarrow 0, \quad \text{as } \epsilon(n) \downarrow 0.$$

Thus  $W_1(\lambda)$  converges to a Gaussian process over  $\Lambda_p$ .

For  $W_2(\lambda)$ , we use a different approach. Although  $W_2$  is a sum of dependent processes, it can be re-written as a sum of independent stochastic processes:

$$\begin{aligned} W_2(\lambda) &= 2\sqrt{n} \sum_{j=1}^p (\lambda_j - 1) \theta_j (\widehat{\theta}_j - \theta) \\ &= \frac{2}{\sqrt{n}} \sum_{j=1}^p (\lambda_j - 1) \sum_{i=1}^n \theta_j (\phi_j(Y_i) - \theta) \\ &= \frac{2}{\sqrt{n}} \sum_{i=1}^n (\theta(\lambda) - \theta)^T (X_i - \theta) \\ &= 2 \sum_{i=1}^n T_i(\lambda) \end{aligned}$$

where  $T_i(\lambda) \equiv \frac{1}{\sqrt{n}}(\theta(\lambda) - \theta)^T(X_i - \theta)$ .

Hence we can use the classical empirical process theory to show stochastic convergence of  $W_2$ . In other words, we only need to show that conditions (34), (35) and (36) are satisfied.

A simple calculus argument shows that

$$\begin{aligned} \|T_i(\lambda)\|_{\Lambda_p}^2 &= \sup_{\Lambda_p} \left( \sum_{j=1}^p (\lambda_j - 1) \theta_j (\phi_j(Y_i) - \theta_j) \right)^2 \\ &\leq \left( \sup_{\Lambda_p} \sum_{j=1}^p \left| (\lambda_j - 1) \theta_j (\phi_j(Y_i) - \theta_j) \right| \right)^2 \\ &\leq \left( \sum_{j=1}^p \left| \theta_j (\phi_j(Y_i) - \theta_j) \right| \right)^2. \end{aligned}$$

Define  $U_i = \sum_{j=1}^p U_{ij}$  where  $U_{ij} \equiv \theta_j (\phi_j(Y_i) - \theta_j)$  for  $i = 1, \dots, n$ .

Owing to  $|\phi_j(Y_i) - \theta_j| \leq 2\sqrt{2}$  for all  $j$ , for  $m \geq 1$ ,

$$\lim_n \mathbb{E} \left( \|T_i(\lambda)\|_{\Lambda_p}^2 \right)^m \leq \lim_n \mathbb{E} (U_i)^{2m} \leq \lim_n \left( 2\sqrt{2} \sum_{j=1}^p |\theta_j| \right)^{2m} < \infty.$$

Furthermore,

$$\begin{aligned} \mathbb{E} \left( \sum_{i=1}^n I\{\|T_i(\lambda)\|_{\Lambda_p}^2 > u\} \|T_i(\lambda)\|_{\Lambda_p}^2 \right) &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left( I\{U_i^2 > nu\} U_i^2 \right) \\ &\leq \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{E} (I\{U_i^2 > nu\}) \mathbb{E} (U_i^4) \right\}^{1/2} \\ &\leq \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{P}(U_i^2 > nu) \mathbb{E} (U_i^4) \right\}^{1/2} \\ &\leq \frac{1}{n\sqrt{nu}} \sum_{i=1}^n [\mathbb{E} (U_i^2) \mathbb{E} (U_i^4)]^{1/2} \\ &= O \left( \frac{1}{\sqrt{n}} \right). \end{aligned}$$

Finally, we need to show that the condition (36) is satisfied for  $W_2(\lambda)$ .

Define  $d_{\widehat{Q}}(s, t) = \max_i d_{Q_i}(s, t)$  where  $d_{Q_i}(s, t) \equiv \left( \int (s - t)^2 dQ_i \right)^{1/2}$  and

$$d_{Q_i}(\cdot) \equiv \begin{cases} \frac{\sum_{j=1}^p \delta_{t_j}(\cdot) U_{ij}}{\sum_{j=1}^n U_{ij}} & : \sum_{j=1}^p U_{ij} > 0 \quad \text{a.e.} \\ 0 & : \text{otherwise.} \end{cases}$$

For arbitrary  $s, t \in \Lambda_p$ ,

$$\begin{aligned} \widehat{\rho}_{W_2}(s, t)^2 &= \frac{4}{n} \sum_{i=1}^n \left( \sum_{j=1}^p (s_j - t_j) \theta_j (\phi_j(Y_i) - \theta_j) \right)^2 \\ &\leq \frac{4}{n} \sum_{i=1}^n \left( \sum_{j=1}^p |s_j - t_j| U_{ij} \right)^2 \\ &= \frac{4}{n} \sum_{i=1}^n \left( d'_{Q_i}(s, t) \right)^2 \left( \sum_{j=1}^p U_{ij} \right)^2 \\ &\leq \frac{4}{n} \sum_{i=1}^n \left( d_{Q_i}(s, t) \right)^2 U_i^2 \\ &\leq \left( d_{\widehat{Q}}(s, t) \right)^2 \frac{4}{n} \sum_{i=1}^n U_i^2, \end{aligned}$$

where  $d'_{Q_i}(s, t) \equiv \int |s - t| dQ_i$ .

Using a similar argument in the proof for  $W_1(\lambda)$ ,

$$N(u, \Lambda_p, \widehat{\rho}_{W_2}) \leq N\left(u \left[ \frac{4}{n} \sum_{i=1}^n U_i^2 \right]^{-1/2}, \Lambda_p\right).$$

And

$$\begin{aligned} \int_0^{\epsilon(n)} \sqrt{\log N(u, \Lambda_p, \widehat{\rho}_{W_2})} du &\leq \int_0^{\epsilon(n)} \left\{ \log N\left(u \left[ \frac{4}{n} \sum_{i=1}^n U_i^2 \right]^{-1/2}, \Lambda_p\right) \right\}^{1/2} du \\ &= \left[ \frac{4}{n} \sum_{i=1}^n U_i^2 \right]^{1/2} \int_0^{\epsilon(n) \left[ (4/n) \sum_{i=1}^n U_i^2 \right]^{1/2}} \sqrt{\log N(u, \Lambda_p)} du \\ &= o_p(1). \end{aligned}$$

Consequently,  $W_2(\lambda)$  converges to a Gaussian process over  $\Lambda_p$ .  $\square$

Since,

$$\begin{aligned}
\text{Var}(W_2(\lambda)) &= \text{Var}(T_1(\lambda)) \\
&= (\theta(\lambda) - \theta)^T \Sigma (\theta(\lambda) - \theta) \\
&= \sum_{j=1}^p \theta_j^2 \sigma_j^2 (\lambda_j - 1)^2 + 2 \sum_{1 \leq k < j \leq p} \theta_j \theta_k \sigma_{jk} (\lambda_j - 1)(\lambda_k - 1) \\
&= \tau_2^2(\lambda),
\end{aligned}$$

with Lemma 4.3, one can show that

$$\begin{aligned}
&\left| \mathbb{E} \left[ \exp \left( it \left( \frac{a(\tilde{\lambda})}{\tau_1^2(\tilde{\lambda})} \right)^{1/2} W_1(\tilde{\lambda}) \right) \right] - \exp \left( - \frac{t^2 a(\tilde{\lambda})}{2} \right) \right| = o(1), \\
&\left| \mathbb{E} \left[ \exp \left( it \left( \frac{1 - a(\tilde{\lambda})}{\tau_2^2(\tilde{\lambda})} \right)^{1/2} W_2(\tilde{\lambda}) \right) \right] - \exp \left( - \frac{t^2 (1 - a(\tilde{\lambda}))}{2} \right) \right| = o(1),
\end{aligned}$$

if

$$\liminf_n \inf_{\Theta(m,C)} \tau_1^2(\tilde{\lambda})/a(\tilde{\lambda}) > 0, \quad \liminf_n \inf_{\Theta(m,C)} \tau_2^2(\tilde{\lambda})/(1 - a(\tilde{\lambda})) > 0.$$

Hence, to derive the asymptotic distribution of  $W(\lambda)/\tau(\lambda)$ , it remains to show that conditions (19) and (20) are satisfied.

LEMMA 4.4. *The followings hold:*

$$\begin{aligned}
&\sup_{\Theta(m,C)} \text{Cov}(W_1(\tilde{\lambda}), W_2(\tilde{\lambda})) = o(1), \\
&\liminf_n \inf_{\Theta(m,C)} \left( \frac{\tau_1^2(\tilde{\lambda}) \tau_2^2(\tilde{\lambda})}{a(\tilde{\lambda})(1 - a(\tilde{\lambda}))} \right)^{1/2} = \liminf_n \inf_{\Theta(m,C)} \left( \frac{n \tau^2(\tilde{\lambda})}{p} \right) > 0.
\end{aligned}$$

PROOF. First,

$$\begin{aligned}
\text{Cov}(W_1(\lambda), W_2(\lambda)) &= \frac{2}{\sqrt{p}} \sum_{j=1}^p (\lambda_j - 1)(2\lambda_j - 1) \text{Cov}(W_{1j}, W_{2j}) \\
&\quad + \frac{4}{\sqrt{p}} \sum_{1 \leq k < j \leq p} (\lambda_j - 1)(2\lambda_k - 1) \text{Cov}(W_{1j}, W_{2k}).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{Cov}(W_{1j}, W_{2j}) &= \theta_j \text{Cov}(E_j, E_j^2) \\
&= 2\theta_j^2 \sigma_j^2 + \frac{1}{n} \theta_j P_n(\theta), \\
\text{Cov}(W_{1j}, W_{2k}) &= \theta_j \text{Cov}(E_j, E_k^2) \\
&= 2\theta_j \theta_k \sigma_{jk} + \frac{1}{n} \theta_j P_n(\theta),
\end{aligned}$$

where  $P_n(\theta)$  is a polynomial of  $\theta$ .

Therefore,

$$\limsup_n \sup_{\Theta(m, C)} \text{Cov}(W_1(\lambda), W_2(\lambda)) = o(1).$$

Now,

$$\liminf_n \inf_{\Theta(m, C)} \left( \frac{\tau_1^2(\tilde{\lambda}) \tau_2^2(\tilde{\lambda})}{a(\tilde{\lambda})(1 - a(\tilde{\lambda}))} \right) = \liminf_n \inf_{\Theta(m, C)} \frac{(p/n) \tau_1^2(\lambda) + \tau_2^2(\lambda)}{p/n} > 0.$$

In other words,

$$\liminf_n \inf_{\theta \in \Theta(\alpha, C)} (\tau_1^2(\tilde{\lambda}) + (n/p) \tau_2^2(\tilde{\lambda})) > 0.$$

We now follow the proof of Beran and Dümbgen (1998) Theorem 3.2.

Suppose that  $\tau_1^2 + \tau_2^2$  can be degenerated at  $\lambda = \tilde{\lambda}$ . Recall that  $\tau_2^2(\tilde{\lambda}) = (\theta(\lambda) - \theta) \Sigma(\theta(\lambda) - \theta)$  is 0 if and only if  $\sum_{j=1}^p \theta_j^2 (\tilde{\lambda}_j - 1)^2 = 0$ .

In other words, we assume

$$\liminf_n \inf_{\theta \in \Theta(\alpha, C)} \left( \frac{1}{p} \sum_{j=1}^p \sigma_j^2 \left( \tilde{\lambda}_j - \frac{1}{2} \right)^2 + \frac{n}{p} \sum_{j=1}^p \theta_j^2 (\tilde{\lambda}_j - 1)^2 \right) = 0 \quad (28)$$

and then we derive a contradiction. Now,

$$\begin{aligned}
\frac{n}{p} R_p(\tilde{\lambda}) &= \frac{1}{p} \sum_{j=1}^p \tilde{\lambda}_j^2 \sigma_j^2 + \frac{n}{p} \sum_{j=1}^p \theta_j^2 (1 - \tilde{\lambda}_j)^2 \\
&\geq \frac{1}{p} \sum_{j=1}^p \tilde{\lambda}_j^2 \sigma_j^2 \\
&\geq \frac{1}{4p} \sum_{j=1}^p \sigma_j^2 - \frac{1}{p} \sum_{j=1}^p \left[ \sigma_j^2 \left( \tilde{\lambda}_j - \frac{1}{2} \right)^2 \right]^{1/2}.
\end{aligned}$$



Let  $\lambda_j^c = \tilde{\lambda}_j 1\{\tilde{\lambda}_j \geq 3/4\}$ , then  $\lambda^c = (\lambda_1^c, \dots, \lambda_p^c) \in \Lambda_p$ . Note that  $(\lambda_j^c)^2 \leq 9(\lambda_j - 1/2)^2$  and  $(1 - \lambda_j^c)^2 \leq 16(1 - \lambda_j)^2$ .

Then,

$$\begin{aligned} \frac{n}{p} R_p(\tilde{\lambda}) &\leq \frac{n}{p} R_p(\lambda^c) \\ &= \frac{1}{p} \sum_{j=1}^p (\lambda_j^c)^2 \sigma_j^2 + \frac{n}{p} \sum_{j=1}^p \theta_j^2 (1 - \lambda_j^c)^2 + \frac{1}{n} \\ &\leq \frac{9}{p} \sum_{j=1}^p \left( \tilde{\lambda}_j - \frac{1}{2} \right)^2 \sigma_j^2 + \frac{16n}{p} \sum_{j=1}^p \theta_j^2 (1 - \lambda_j)^2 \\ &= o(1). \end{aligned}$$

These two inequalities contradict the equation (28).  $\square$

Our next task is to show that  $W_k(\hat{\lambda})$  is stochastically very close to  $W_k(\tilde{\lambda})$  for  $k = 1, 2$ .

The theorem below not only plays a key role in the proof of stochastic closeness but also show convergence of loss and risk functions.

**THEOREM 4.1.** *Let  $\hat{\lambda}$  and  $\tilde{\lambda}$  be minimizers of  $L_p(\lambda)$  and  $R_p(\lambda)$  over  $\Lambda_p$ .*

*Then  $\mathbb{E} \left( \sum_{j=1}^p \left( \frac{\sigma_j^2}{n} + \theta_j^2 \right) \left( \hat{\lambda}_j - \tilde{\lambda}_j \right)^2 \right)$  and  $\mathbb{E} \left( \sum_{j=1}^p \hat{\theta}_j^2 \left( \hat{\lambda}_j - \tilde{\lambda}_j \right)^2 \right)$  are bounded by*

$$C \frac{J(\Lambda_p)}{\sqrt{n}} \left[ \left( \frac{1}{n} \sum_{j=1}^p \mathbb{E}(W_{1j}^2) \right)^{1/2} + \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}(U_i^2) \right)^{1/2} + O\left(\frac{p}{n}\right) \right],$$

where  $J(\Lambda_p) = \int_0^1 \sqrt{\log N(u, \Lambda_p)} du$ .

Furthermore,

$$|L_p(\hat{\lambda}) - R_p(\tilde{\lambda})| = o_p(1), \quad \text{and} \quad \left| \frac{1}{p} \sum_{j=1}^p (\tilde{\lambda}_j - \hat{\lambda}_j)^2 \right| = o_p(1).$$

**PROOF.** Let  $w_{1j} = \hat{\theta}_j^2$ ,  $w_{2j} = \frac{\sigma_j^2}{n} + \theta_j^2$ ,  $g_{1j} = \left( \hat{\theta}_j^2 - \frac{\hat{\sigma}_j^2}{n} \right) / \hat{\theta}_j^2$  and  $g_{2j} = \theta_j^2 / \left( \frac{\sigma_j^2}{n} + \theta_j^2 \right)$  for  $j = 1, \dots, p$ . Then  $\hat{\lambda} = \lambda_1^o$  and  $\tilde{\lambda} = \lambda_2^o$  where

$$\lambda_i^o \equiv \arg \min_{\lambda \in \Lambda_p} \sum_{j=1}^p [w_{ij}(\lambda_j - g_{ij})^2], \quad \text{for } i = 1, 2.$$

Using the same argument in the proof of Theorem 2.2 in Beran and Dümbgen (1998), one can show that

$$\begin{aligned}
\mathbb{E} \left( \sum_{j=1}^p \left( \frac{\sigma_j^2}{n} + \theta_j^2 \right) (\hat{\lambda}_j - \tilde{\lambda}_j)^2 \right) &\leq \sum_{j=1}^p \mathbb{E} \left| \left( \hat{\theta}_j^2 (\hat{\lambda}_j - g_{1j}) - \left( \frac{\sigma_j^2}{n} + \theta_j^2 \right) (\hat{\lambda}_j - g_{2j}) \right) (\tilde{\lambda}_j - \hat{\lambda}_j) \right| \\
&\leq \frac{1}{\sqrt{n}} \sum_{j=1}^p \mathbb{E} \left| \left( (\hat{\lambda}_j - 1) \left( 2W_{2j} + \frac{W_{1j}}{\sqrt{n}} \right) + \frac{V_j}{n} \right) (\tilde{\lambda}_j - \hat{\lambda}_j) \right| \\
&\leq \frac{4}{\sqrt{n}} \mathbb{E} \left( \sup_{\mathbf{g} \in \mathcal{G}} \left| \sum_{j=1}^p g_j \left( \frac{W_{1j}}{\sqrt{n}} + 2W_{2j} + \frac{V_j}{n} \right) \right| \right).
\end{aligned}$$

where  $\mathcal{G} \equiv \{\mathbf{fg} : \mathbf{f}, \mathbf{g} \in \Lambda_p\}$ .

With the maximal inequality

$$\begin{aligned}
\mathbb{E} \left( \sup_{\mathbf{g} \in \mathcal{G}} \frac{1}{\sqrt{p}} \left| \sum_{j=1}^p g_j W_{1j} \right| \right) &\leq C \mathbb{E} \left( \int_0^{\hat{D}_{W_1}} \sqrt{\log N(u, \mathcal{G}, \hat{\rho}_S)} du \right) \\
&= C \mathbb{E} \left( \frac{1}{p} \sum_{j=1}^p W_{1j}^2 \right)^{1/2} \int_0^1 \sqrt{\log N(u, \mathcal{G})} du \\
&\leq C J(\mathcal{G}) \left( \frac{1}{p} \sum_{j=1}^p \mathbb{E} (W_{1j}^2) \right)^{1/2}, \\
\mathbb{E} \left( \sup_{\mathbf{g} \in \mathcal{G}} \left| \sum_{j=1}^p g_j W_{2j} \right| \right) &\leq C \mathbb{E} \left( \int_0^{\hat{D}_T} \sqrt{\log N(u, \mathcal{G}, \hat{\rho}_T)} du \right) \\
&= C \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n U_i^2 \right)^{1/2} \int_0^1 \sqrt{\log N(u, \mathcal{G})} du \\
&\leq C J(\mathcal{G}) \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E} (U_i^2) \right)^{1/2}.
\end{aligned}$$

Since  $N(u, \mathcal{G}) \leq N(u/2, \Lambda_p)^2$  for all  $u > 0$ , it is simple to show that

$$J(\mathcal{G}) \leq 4J(\Lambda_p) = O(1).$$

Furthermore,

$$\mathbb{E} \left( \sup_{\mathbf{g} \in \mathcal{G}} \frac{1}{p} \left| \sum_{j=1}^p g_j V_j \right| \right) \leq \mathbb{E} \left( \sup_{\mathbf{g} \in \mathcal{G}} \sum_{j=1}^p \frac{1}{p} |g_j V_j| \right)$$

$$\begin{aligned}
&\leq \frac{1}{p} \mathbb{E} \left( \sum_{j=1}^p |V_j| \right) \\
&\leq \frac{1}{p} \sum_{j=1}^p \left( \mathbb{E} (V_j^2) \right)^{1/2} \\
&= O(1).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E} \left( \sum_{j=1}^p \left( \frac{\sigma_j^2}{n} + \theta_j^2 \right) (\hat{\lambda}_j - \tilde{\lambda}_j)^2 \right) &\leq C \frac{J(\Lambda_p)}{\sqrt{n}} \left[ \left( \frac{1}{n} \sum_{j=1}^p \mathbb{E} (W_{1j}^2) \right)^{1/2} + \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E} (U_i^2) \right)^{1/2} + O\left(\frac{p}{n}\right) \right] \\
&= O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E} \left( \sum_{j=1}^p \hat{\theta}_j^2 (\hat{\lambda}_j - \tilde{\lambda}_j)^2 \right) &\leq \frac{1}{\sqrt{n}} \sum_{j=1}^p \mathbb{E} \left( \left[ \left( 1 - \tilde{\lambda}_j \right) \left( 2W_{2j} + \frac{W_{1j}}{\sqrt{n}} \right) - \frac{V_j}{n} \right] (\tilde{\lambda}_j - \hat{\lambda}_j) \right) \\
&\leq \frac{4}{\sqrt{n}} \mathbb{E} \left( \sup_{\mathbf{g} \in \mathcal{G}} \left| \sum_{j=1}^p \lambda_j \left( \frac{W_{1j}}{\sqrt{n}} + 2W_{2j} + \frac{V_j}{n} \right) \right| \right) \\
&= C \frac{J(\Lambda_p)}{\sqrt{n}} \left[ \left( \frac{1}{n} \sum_{j=1}^p \mathbb{E} (W_{1j}^2) \right)^{1/2} + \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E} (U_i^2) \right)^{1/2} + O\left(\frac{p}{n}\right) \right] \\
&= O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Now,

$$\frac{1}{p} \sum_{j=1}^p \sigma_j^2 (\tilde{\lambda}_j - \hat{\lambda}_j)^2 \leq \frac{1}{p} \left| \sum_{j=1}^p (\sigma_j^2 - E_j^2) (\tilde{\lambda}_j - \hat{\lambda}_j)^2 \right| + \frac{1}{p} \left| \sum_{j=1}^p E_j^2 (\tilde{\lambda}_j - \hat{\lambda}_j)^2 \right|.$$

One can show the following with the maximal inequality :

$$\left| \frac{1}{p} \sum_{j=1}^p (\sigma_j^2 - E_j^2) (\hat{\lambda}_j - \tilde{\lambda}_j)^2 \right| \leq \frac{4}{p} \sup_{g \in \mathcal{G}} \left| \sum_{j=1}^p g_j W_{1j} \right| = O\left(\frac{1}{\sqrt{p}}\right).$$

For the second term of the right hand side, one can show

$$\frac{1}{p} \sum_{j=1}^p E_j^2 (\hat{\lambda}_j - \tilde{\lambda}_j)^2 \leq \frac{1}{p} \sum_{j=1}^p E_j^2.$$

Furthermore,

$$\text{Var} \left( \sum_{j=1}^p E_j^2 \right) = \sum_{j=1}^p \mathbb{E} (E_j^4) + 2 \sum_{j < k} \text{Cov} (E_j^2, E_k^2) = O(p)$$

Therefore,

$$\frac{1}{p} \sum_{j=1}^p (\tilde{\lambda}_j - \hat{\lambda}_j)^2 = O_p \left( \frac{1}{\sqrt{p}} \right),$$

which follows immediately from

$$\frac{1}{p} (\min_j \sigma_j^2) \sum_{j=1}^p (\tilde{\lambda}_j - \hat{\lambda}_j)^2 \leq \frac{1}{p} \sum_{j=1}^p \sigma_j^2 (\tilde{\lambda}_j - \hat{\lambda}_j)^2 = O_p \left( \frac{1}{\sqrt{p}} \right).$$

Finally, we show convergence of the loss and the risk functions.

By the triangle inequality

$$|L_p(\hat{\lambda}) - R_p(\hat{\lambda})| \leq |L_p(\hat{\lambda}) - R_p(\tilde{\lambda})| + |R_p(\hat{\lambda}) - R_p(\tilde{\lambda})|.$$

For the first term in the above equation,

$$\begin{aligned} |L_p(\hat{\lambda}) - R_p(\hat{\lambda})| &\leq \frac{1}{\sqrt{n}} \left| \sum_{j=1}^p \hat{\lambda}_j^2 \frac{W_{1j}}{\sqrt{n}} \right| + \frac{2}{\sqrt{n}} \left| \sum_{j=1}^p \hat{\lambda}_j (1 - \hat{\lambda}_j) W_{2j} \right| \\ &\leq \frac{\sqrt{p}}{n} \sup_{\mathbf{g} \in \mathcal{G}} \left| \sum_{j=1}^p g_j \frac{W_{1j}}{\sqrt{p}} \right| + \frac{2}{\sqrt{n}} \sup_{\mathbf{g} \in \mathcal{G}} \left| \sum_{j=1}^p g_j W_{2j} \right| \\ &= O_p \left( \frac{\sqrt{p}}{n} \right) + O_p \left( \frac{1}{\sqrt{n}} \right) = o_p(1). \end{aligned}$$

On the other hand,

$$\left| R_p(\hat{\lambda}) - R_p(\tilde{\lambda}) \right| \leq \left| \frac{1}{n} \sum_{j=1}^p \sigma_j^4 (\hat{\lambda}_j^2 - \tilde{\lambda}_j^2) \right| + \left| \sum_{j=1}^n \theta_j^2 \left( (1 - \hat{\lambda}_j)^2 - (1 - \tilde{\lambda}_j)^2 \right) \right|.$$

It is straightforward to show that the first term is  $O_p(p/n)$ .

Then with the Cauchy-Schwartz inequality, one can show that

$$\begin{aligned}
\sum_{j=1}^p \theta_j^2 ((1 - \widehat{\lambda}_j)^2 - (1 - \widetilde{\lambda}_j)^2) &= \sum_{j=1}^p \theta_j^2 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 + 2 \sum_{j=1}^p \theta_j (1 - \widehat{\lambda}_j) (\widehat{\lambda}_j - \widetilde{\lambda}_j) \\
&\leq 2 \left( \sum_{j=1}^p \theta_j^2 (1 - \widehat{\lambda}_j)^2 \right)^{1/2} \left( \sum_{j=1}^p \theta_j^2 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 \right)^{1/2} \\
&\quad + \sum_{j=1}^p \theta_j^2 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 \\
&= o_p(1),
\end{aligned}$$

since  $\sum_{j=1}^p \theta_j^2 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 = o_p(1)$  which follows immediately from

$$\sum_{j=1}^p \left( \theta_j^2 + \frac{\widehat{\sigma}_j^2}{n} \right) (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 = o_p(1).$$

Therefore,  $|R_p(\widehat{\lambda}) - R_p(\widetilde{\lambda})| = o_p(1)$  which completes the proof.  $\square$

With Theorem 4.1, we can show stochastic closeness of  $W_k(\widetilde{\lambda})$  and  $W_k(\widehat{\lambda})$  for  $k = 1, 2$ .

LEMMA 4.5. *For  $k = 1, 2$ ,*

$$\sup_{\Theta(m, C)} (W_k(\widehat{\lambda}) - W_k(\widetilde{\lambda})) = o_p(1).$$

PROOF. By stochastic equicontinuity, it suffices to show that  $\rho_{W_1}(\widehat{\lambda}, \widetilde{\lambda})^2$  and  $\rho_{W_2}(\widehat{\lambda}, \widetilde{\lambda})^2$  are  $o_p(1)$ . It is straightforward to show that

$$\begin{aligned}
\rho_{W_1}(\widehat{\lambda}, \widetilde{\lambda})^2 &= \frac{1}{p} \left[ \mathbb{E} \left( \widehat{\rho}_{W_1}(\lambda, \widetilde{\lambda})^2 \right) \right]_{\lambda=\widehat{\lambda}} \\
&= \mathbb{E} \left[ \sum_{j=1}^p (\lambda_j - \widetilde{\lambda}_j)^2 W_{1j}^2 \right]_{\lambda=\widehat{\lambda}} \\
&= \frac{2}{p} \sum_{j=1}^p \sigma_j^4 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 + o_p(1)
\end{aligned}$$

$$\begin{aligned}
&= o_p(1), \\
\rho_{W_2}(\widehat{\lambda}, \widetilde{\lambda})^2 &= \left[ \mathbb{E} \left( \widehat{\rho}_{W_2}(\lambda, \widetilde{\lambda})^2 \right) \right]_{\lambda=\widehat{\lambda}} \\
&= 4\mathbb{E} \left[ \sum_{j=1}^p (\lambda_j - \widetilde{\lambda}_j) \theta_j (\phi_j(Y_i) - \theta_j) \right]_{\lambda=\widehat{\lambda}}^2 \\
&= 4 \sum_{j=1}^p \theta_j^2 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 \sigma_j^2 + 8 \sum_{j < k}^p \theta_j \theta_k (\widehat{\lambda}_j - \widetilde{\lambda}_j) (\widehat{\lambda}_k - \widetilde{\lambda}_k) \sigma_{jk} \\
&= 4(\theta(\widehat{\lambda}) - \theta(\widetilde{\lambda}))^T \Sigma (\theta(\widehat{\lambda}) - \theta(\widetilde{\lambda})) \\
&= o_p(1).
\end{aligned}$$

The last equality follows from the fact that  $\Sigma$  is a positive definite therefore it can be 0 if and only if  $\sum_{j=1}^p \theta_j^2 (\widehat{\lambda}_j - \widetilde{\lambda}_j)^2 = 0$  which is shown in Theorem 4.1.  $\square$

Recall that

$$W(\widehat{\lambda})/\tau(\widetilde{\lambda}) = (a(\widetilde{\lambda}))^{1/2} (W_1(\widehat{\lambda})/\tau_1(\widetilde{\lambda})) + (1 - a(\widetilde{\lambda}))^{1/2} (W_2(\widehat{\lambda})/\tau_2(\widetilde{\lambda})),$$

where  $a(\lambda) = (p/n)\tau_1^2(\lambda)/\tau^2(\lambda)$ .

We already showed that  $W(\widehat{\lambda})/\tau(\widetilde{\lambda})$  converges to a standard Normal. With an application of Lemma 4.5, one can show that

$$\sup_{\Theta(m, C)} |W(\widehat{\lambda})/\tau(\widetilde{\lambda}) - W(\widetilde{\lambda})/\tau(\widetilde{\lambda})| = o_p(1).$$

since  $a(\widetilde{\lambda})$  is bounded.

Now we prove Theorem 3.2.

PROOF. First, we show that  $\widehat{\tau}_k^2(\widetilde{\lambda})$  is a consistent estimator for  $k = 1, 2$ .

From the proof of  $\text{Var}(V(\lambda)) = O(1)$ , one can obtain

$$\frac{1}{p} \sum_{j=1}^p \widehat{\sigma}_j^2 = \frac{1}{p} \sum_{j=1}^p \sigma_j^2 + o_p(1). \quad (29)$$

By the triangle inequality,

$$|\widehat{\tau}_1^2(\widehat{\lambda}) - \tau_1^2(\widetilde{\lambda})| \leq \frac{2}{p} \sum_{j=1}^p |\widehat{\sigma}_j^4 (2\widehat{\lambda}_j - 1)^2 - \sigma_j^4 (2\widetilde{\lambda}_j - 1)^2|$$

$$+\frac{2}{p}\sum_{j=1}^p|\sigma_j^4(2\widehat{\lambda}_j-1)^2-\sigma_j^4(2\widetilde{\lambda}_j-1)^2|$$

While the equation (29) ensures convergence of the first term in right hand side to 0, applying the Cauchy-Schwartz inequality with  $\frac{1}{p}\sum_{j=1}^p(\widehat{\lambda}-\widetilde{\lambda})^2=o_p(1)$  shows that the second term converges to 0.

Now,

$$\begin{aligned} |\widehat{\tau}_2^2(\widehat{\lambda})-\tau_2^2(\widetilde{\lambda})| &\leq \sum_{j=1}^p\left[\left(\widehat{\theta}_j^2-\frac{\widehat{\sigma}_j^2}{n}\right)\widehat{\sigma}_j^2(\widehat{\lambda}_j^2-1)-\theta_j^2\sigma_j^2(\widetilde{\lambda}_j^2-1)\right] \\ &\quad +2\sum\sum_{1\leq k<j\leq p}\left(\widehat{\theta}_j\widehat{\theta}_k(\widehat{\lambda}_j-1)(\widehat{\lambda}_k-1)\widehat{\sigma}_{jk}-\theta_j\theta_k(\widetilde{\lambda}_j-1)(\widetilde{\lambda}_k-1)\sigma_{jk}\right). \end{aligned}$$

Combining the maximal inequality and results of Theorem 4.1, one can show

$$|S_p(\widehat{\lambda})-R_p(\widetilde{\lambda})|=o_p(1).$$

Consequently,

$$\begin{aligned} \sum_{j=1}^p\left(\widehat{\theta}_j^2-\frac{\widehat{\sigma}_j^2}{n}\right)(\widehat{\lambda}_j-1)^2 &= S_p(\widehat{\lambda})-\frac{1}{n}\sum_{j=1}^p\widehat{\lambda}_j^2\widehat{\sigma}_j^2 \\ &= R_p(\widetilde{\lambda})-\frac{1}{n}\sum_{j=1}^p\widetilde{\lambda}_j^2\sigma_j^2+o_p(1) \\ &= \sum_{j=1}^p\theta_j^2(\widetilde{\lambda}_j-1)^2+o_p(1). \end{aligned}$$

Applying the above result with (29) yields

$$\sum_{j=1}^p\left|\left(\widehat{\theta}_j^2-\frac{\widehat{\sigma}_j^2}{n}\right)\widehat{\sigma}_j^2(\widehat{\lambda}_j^2-1)-\theta_j^2\sigma_j^2(\widetilde{\lambda}_j^2-1)\right|=o_p(1).$$

For the second term,

$$\left|\sum\sum_{1\leq k<j\leq p}\left(\widehat{\theta}_j\widehat{\theta}_k(\widehat{\lambda}_j-1)(\widehat{\lambda}_k-1)\widehat{\sigma}_{jk}-\theta_j\theta_k(\widetilde{\lambda}_j-1)(\widetilde{\lambda}_k-1)\sigma_{jk}\right)\right|,$$

one can show it is bounded by

$$\begin{aligned} & \left| \sum_{k < j} (\hat{\theta}_j \hat{\theta}_k \hat{\sigma}_{jk} - \theta_j \theta_k \sigma_{jk}) (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) \right| \\ & + \left| \sum_{k < j} \theta_j \theta_k \sigma_{jk} \left( (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) - (\tilde{\lambda}_j - 1) (\tilde{\lambda}_k - 1) \right) \right|. \end{aligned} \quad (30)$$

The first term in (30) can be bounded by

$$\begin{aligned} \left| \sum_{k < j} (\hat{\theta}_j \hat{\theta}_k \hat{\sigma}_{jk} - \theta_j \theta_k \sigma_{jk}) (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) \right| & \leq \left| \sum_{k < j} (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) (\hat{\theta}_j \hat{\theta}_k - \theta_j \theta_k) \hat{\sigma}_{jk} \right| \\ & + \sum_{k < j} |\theta_j \theta_k| |\hat{\sigma}_{jk} - \sigma_{jk}|. \end{aligned}$$

Since  $\hat{\sigma}_{jk}$  is a consistent estimator of  $\sigma_{jk}$  and  $\sum_{j,k} |\theta_j \theta_k|$  is bounded, one can show

$$\sum_{k < j} |\theta_j \theta_k| |\hat{\sigma}_{jk} - \sigma_{jk}| = O_p \left( \frac{1}{\sqrt{n}} \right).$$

Furthermore,

$$\begin{aligned} & \left| \sum_{1 \leq k < j \leq p} (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) (\hat{\theta}_j \hat{\theta}_k - \theta_j \theta_k) \hat{\sigma}_{jk} \right| \\ & \leq \left| \sum_{1 \leq k < j \leq p} (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) (\hat{\theta}_j - \theta_j) \hat{\theta}_k \hat{\sigma}_{jk} \right| + \left| \sum_{1 \leq k < j \leq p} (\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) (\hat{\theta}_k - \theta_k) \theta_j \hat{\sigma}_{jk} \right| \\ & \leq \sum_{1 \leq k < j \leq p} |(\hat{\theta}_j - \theta_j) \hat{\theta}_k \hat{\sigma}_{jk}| + \sum_{1 \leq k < j \leq p} |(\hat{\theta}_k - \theta_k) \theta_j \hat{\sigma}_{jk}| \\ & \leq C_1 \sum_{j=1}^p |\hat{\theta}_j - \theta_j| + C_2 \sum_{k=1}^p |\hat{\theta}_k - \theta_k| \\ & = O_p \left( \frac{p}{\sqrt{n}} \right). \end{aligned}$$

Now it remains to show convergence of the second term in (30) to 0.

$$\sum_{k < j} \theta_j \theta_k \sigma_{jk} [(\hat{\lambda}_j - 1) (\hat{\lambda}_k - 1) - (\tilde{\lambda}_j - 1) (\tilde{\lambda}_k - 1)]$$



$$= \sum_{k < j} \theta_j \theta_k \sigma_{jk} \left( (\hat{\lambda}_j - \tilde{\lambda}_j)(\hat{\lambda}_k - \tilde{\lambda}_k) - (\tilde{\lambda}_k + 1)(\hat{\lambda}_j - \tilde{\lambda}_j) - (\tilde{\lambda}_j + 1)(\hat{\lambda}_k - \tilde{\lambda}_k) \right).$$

Our first goal is to show that

$$\left| \sum_{j=2}^p \sum_{k=1}^j \theta_j \theta_k \sigma_{jk} (\tilde{\lambda}_k + 1)(\hat{\lambda}_j - \tilde{\lambda}_j) \right| = o_p(1).$$

Let  $C_j(\theta) = \sum_{k=1}^j \sigma_{jk} \theta_k (\tilde{\lambda}_k + 1)$ . Since  $\sigma_{jk}$ 's are bounded,  $C_j(\theta) = O(1)$ . Then,

$$\begin{aligned} \left| \sum_{j=2}^p \sum_{k=1}^j \theta_j \theta_k \sigma_{jk} (\tilde{\lambda}_k + 1)(\hat{\lambda}_j - \tilde{\lambda}_j) \right| &\leq \left| \sum_{j=2}^p \theta_j (\hat{\lambda}_j - \tilde{\lambda}_j) C_j(\theta) \right| \\ &\leq \sum_{j=2}^p \theta_j^2 (\hat{\lambda}_j - \tilde{\lambda}_j)^2 \sum_{j=2}^p C_j(\theta)^2 \\ &= O_p\left(\frac{p}{\sqrt{n}}\right) \\ &= o_p(1). \end{aligned}$$

Applying the same argument for the rest of terms provides convergence of each term to 0. Therefore  $\hat{\tau}_2^2(\hat{\lambda})$  is a uniformly consistent estimator of  $\tau_2^2(\tilde{\lambda})$ . We already showed in Lemma 4.4 that the pivot process does not approach a degenerate distribution.

Finally,

$$\begin{aligned} P(\theta \in \mathcal{D}_p) &= P\left(\sum_{j=1}^p (\theta_j - \hat{\lambda}_j \hat{\theta}_j)^2 \leq \frac{z_\alpha \hat{\tau}(\hat{\lambda})}{\sqrt{n}} + S_p(\hat{\lambda})\right) \\ &= P\left(\frac{\sqrt{n}(L_p(\hat{\lambda}) - S_p(\hat{\lambda}))}{\hat{\tau}(\hat{\lambda})} \leq z_\alpha\right) \rightarrow 1 - \alpha. \end{aligned}$$

Therefore,  $\mathcal{D}_p$  is a uniform asymptotic  $1-\alpha$  confidence set for  $\theta$ .  $\square$

## 5 Numerical Examples

In this section, we apply the REACT density estimator to some examples. As test cases, we use the uniform distribution and the mixture Normal distribu-

| sample size | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
|-------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1000        | 95 | 94 | 84 | 82 | 90 | 92 | 86 | 92 | 88 | 81 | 89 | 91 | 91 | 85 |
| 5000        | 98 | 96 | 90 | 81 | 80 | 83 | 97 | 92 | 89 | 94 | 84 | 85 | 88 | 77 |
| 10000       | 98 | 93 | 92 | 85 | 86 | 82 | 95 | 95 | 98 | 96 | 86 | 86 | 90 | 81 |
| 100000      | 98 | 96 | 90 | 81 | 80 | 83 | 97 | 92 | 89 | 94 | 85 | 85 | 88 | 77 |

Table 1: 95% Confidence band coverage for each densities with different sample size

tion examples from Marron and Wand (1992). Some of the mixture models are modified for simplicity.

We use a nested subset selection class as  $\Lambda_p$  instead of monotone class for simplicity. In other words,  $\lambda$  has a form of  $\lambda = (1, \dots, 1, 0, \dots, 0)$ . Then the corresponding orthogonal density estimator is

$$\hat{f}(y) = 1 + \sum_{j=1}^{\hat{J}} \hat{\theta}_j \phi_j(y),$$

where  $\hat{J}$  minimizes the risk function estimator

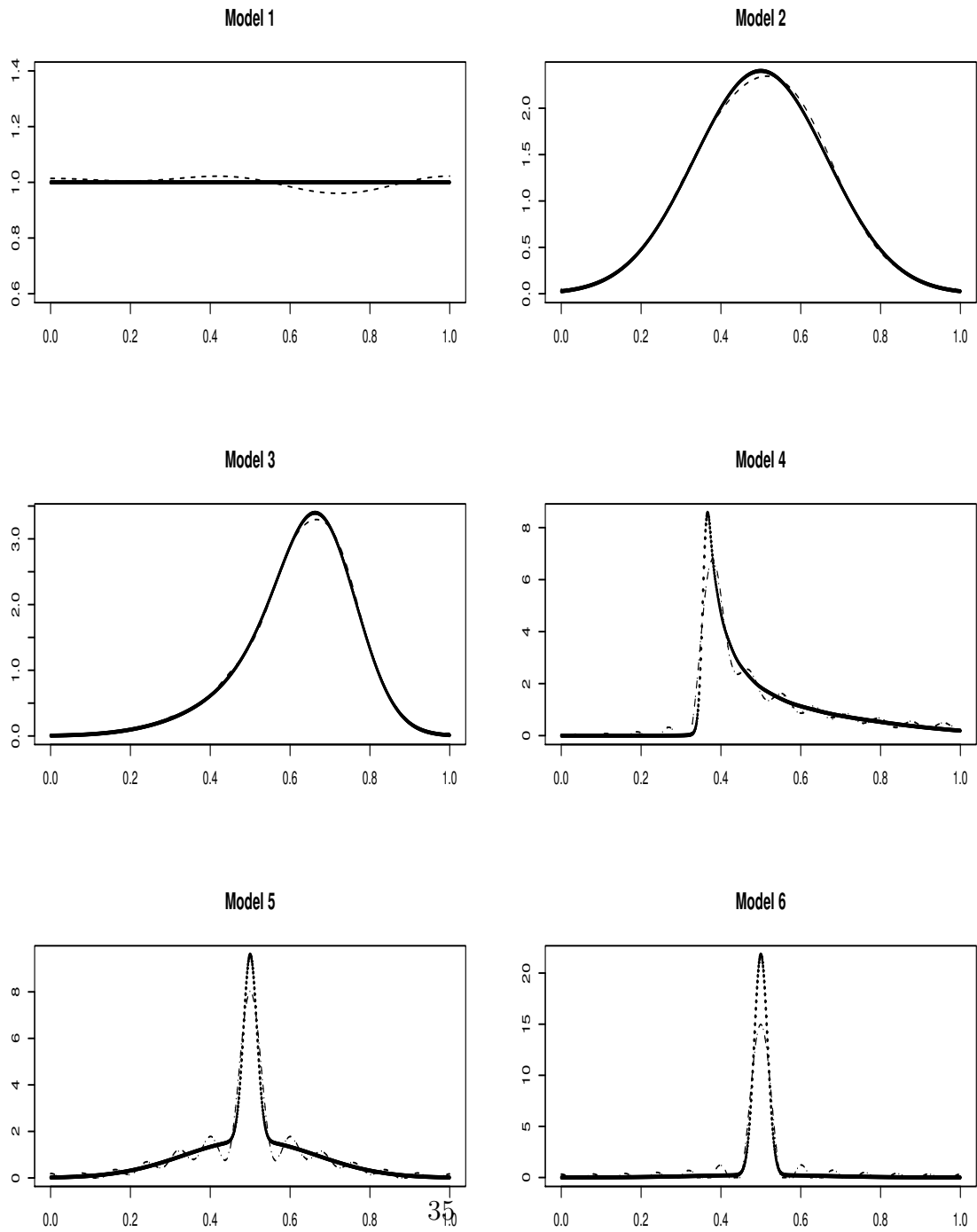
$$S_p(J) = \sum_{j=1}^J \frac{\hat{\sigma}_j^2}{n} + \sum_{j=J+1}^p \left( \theta_j^2 - \frac{\hat{\sigma}_j^2}{n} \right).$$

Figure 1 show the true densities, the projection densities and the density estimators in each model.

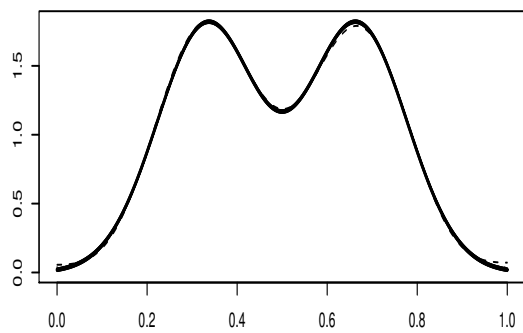
Table 1 reports the coverage of confidence balls for the coefficients of uniform and each mixture Normal with four different sample sizes. Since we estimate the projection  $f_p$  of the true density, the coverage for projection densities is presented.

The performance of the confidence set is uneven. However, the reader should bear in mind that many of the test cases we constructed by Marron and Wand to be difficult. Thus, it is not surprising that the coverage is low in some cases.

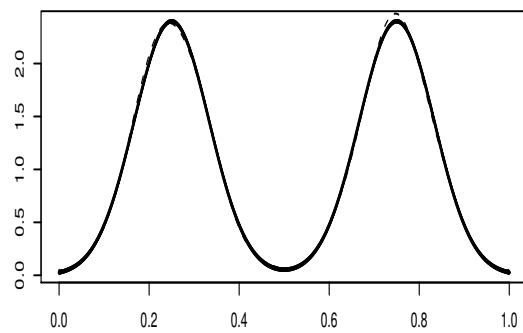
Figure 1: : Densities(solid), Projection densities(dotted) and Density Estimators(dashed) with sample size 20000



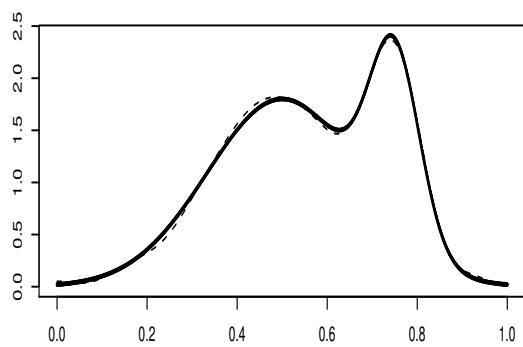
Model 7



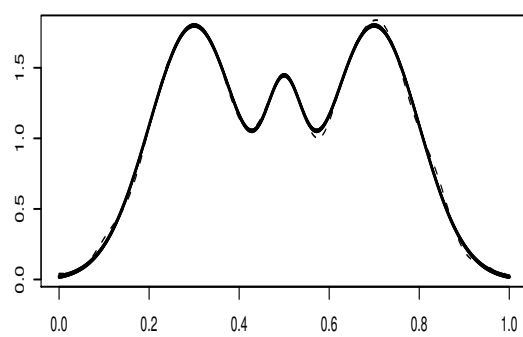
Model 8



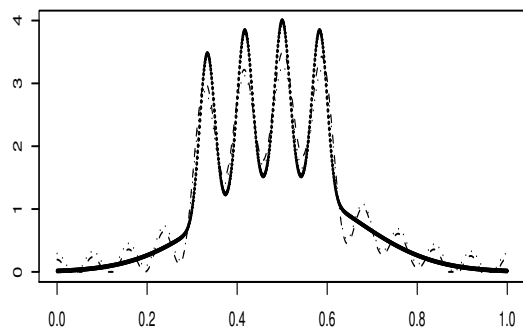
Model 9



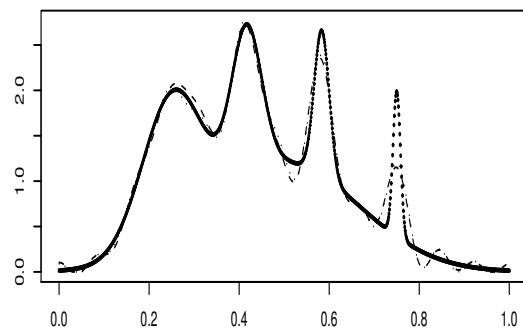
Model 10

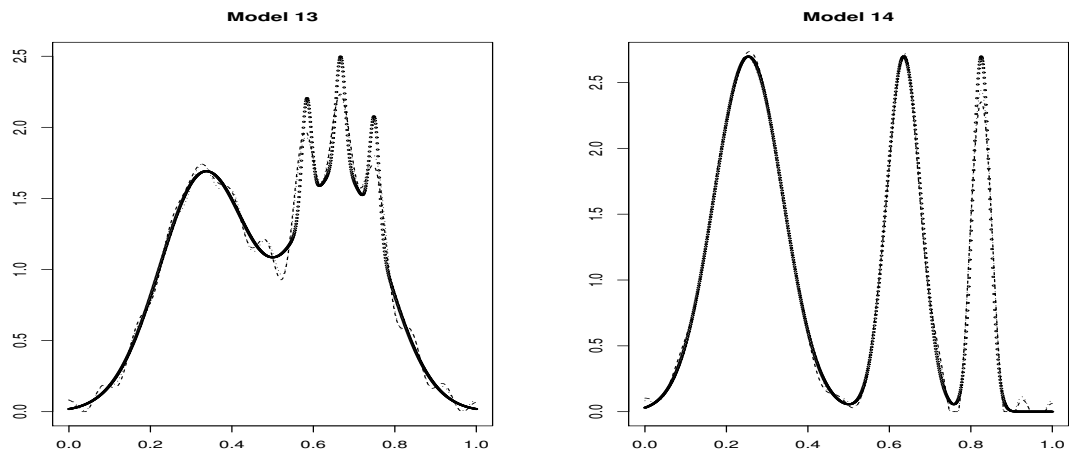


Model 11



Model 12





## 6 Discussion

We have shown how to construct a nonparametric confidence ball for a density function. To our knowledge, this is the first such construction.

An important area for future research is to see if the condition  $p = o(n^{1/3})$  can be weakened although we note that the same condition appears in Nussbaum (1996).

Our preliminary numerical investigations show that the coverage is good in certain cases but can be poor in other cases. Diagnosing and improving these cases of undercoverage remains a challenge.

## Appendix: Empirical Process Theory for the Dependent Case

Let  $S = \sum_{i=1}^n S_i$  where  $S_1, S_2, \dots, S_n$  are stochastic processes on an index set  $\mathcal{T}$  from probability space  $(\Omega, \mathcal{A}, P)$  with norm  $\|x\|_{\mathcal{T}} \equiv \sup_{t \in \mathcal{T}} |x(t)|$ .

Suppose the process  $S$  has continuous sample paths with respect to some

metric  $d$  on  $\mathcal{T}$ . Define the covering number

$$N(u, \mathcal{T}, d_Q) \equiv \min\{\#\mathcal{T}_0 : \mathcal{T}_0 \subset \mathcal{T}, \inf_{t_0 \in \mathcal{T}_0} d_Q(t_0, t) \leq u \quad \forall t \in \mathcal{T}\},$$

and further define the uniform covering number

$$N(u, \mathcal{T}) \equiv \sup_Q N(u, \mathcal{T}, d_Q),$$

where  $d_Q(s, t)^2 \equiv \int (s - t)^2 dQ$  and  $Q$  varies over all probability measures.

Theorem 6.1 and 6.2 are modified versions of the maximal inequality and the functional central limit theorem for sums of dependent stochastic processes.

**THEOREM 6.1.** *Suppose that  $S(t_0) \equiv 0$  for some  $t_0 \in \mathcal{T}$  and that all the finite dimensional distributions of  $S^o \equiv S - \mathbf{E}(S)$  have Gaussian limits. Suppose that*

$$\sup_{s, t \in \mathcal{T}} \sum_{1 \leq i \neq j \leq n} \left| \text{Cov}(S_i^o(s), S_j^o(t)) \right| = o(1). \quad (31)$$

Then,

$$\mathbf{E}(\|S - \mathbf{E}(S)\|_{\mathcal{T}}) \leq C \mathbf{E} \int_0^{\widehat{D}_S} \sqrt{\log N(u, \mathcal{T}, \widehat{\rho}_S)} du,$$

where  $\widehat{D}_S \equiv \sup_{t \in \mathcal{T}} \widehat{\rho}_S(t, t_0)$ .

**PROOF.** For any finite subset  $\{t_1, \dots, t_m\}$  of  $\mathcal{T}$ , the condition (31) implies

$$\left( \mathbf{E}[S^o(t) - S^o(s)]^2 \right)^{1/2} \leq d(s, t) + o(1) \quad \text{for all } s, t \in \mathcal{T}. \quad (32)$$

Since the finite dimensional distribution of  $S^o$  has a Gaussian limit,  $S^o(t_k)$  converges to a Normal for  $k = 1, \dots, m$ . In other words,

$$\left| \mathbf{E} \Psi(S^o(t_k)) - \mathbf{E} \Psi(\eta_k Z_k) \right| \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

where  $\Psi(x) = \exp\left(\frac{x^2}{4\eta^2}\right)$ ,  $\eta_k^2 \equiv \mathbf{E}[S^o(t_k)]^2$ ,  $\eta^2 \equiv \max_{k \leq m} \eta_k^2$  and  $Z_k$ 's are standard Normal random variables.

With the Jensen's inequality, one can show

$$\begin{aligned}
\exp \left( \mathbb{E} \left( \max_{k \leq m} \frac{|S^o(t_k)|^2}{4\eta^2} \right) \right) &= \Psi \left( \mathbb{E} \left( \max_{k \leq m} |S^o(t_k)| \right) \right) \\
&\leq \mathbb{E} \left( \max_{k \leq m} \Psi(|S^o(t_k)|) \right) \\
&\leq \mathbb{E} \left( \sum_{k=1}^m \Psi(|S^o(t_k)|) \right) \\
&\leq \sum_{k=1}^m \mathbb{E} \left( \Psi(|\eta_k Z_k|) \right) + o(m) \\
&\leq \sqrt{2}m + o(m) = O(m).
\end{aligned}$$

Consequently,

$$\left( \mathbb{E} \left( \max_{k \leq m} |S^o(t_k)|^2 \right) \right)^{1/2} \leq C \sqrt{\log m} \max_{k \leq m} \left( \mathbb{E} [S^o(t_k)]^2 \right)^{1/2} + o(1). \quad (33)$$

Using the standard chaining method with the inequalities (32) and (33), it follows that

$$\left( \mathbb{E} \|S^o\|_{\mathcal{T}}^2 \right)^{1/2} \leq \left( \mathbb{E} [S^o(t_o)]^2 \right)^{1/2} + C \sum_{i=0}^{\infty} \mathbb{E} \left( \frac{\hat{D}_s}{2^i} \sqrt{\log N(\delta/2^{i+1}, \mathcal{T}, \hat{\rho}_S)} \right).$$

Let  $\delta_i = \hat{D}_s/2^i$ . Then

$$\mathbb{E} \left( \delta_i \sqrt{\log N(\delta_i, \mathcal{T}, \hat{\rho}_S)} \right) \leq 4 \mathbb{E} \left( \int_{\delta_{i+2}}^{\delta_{i+1}} \sqrt{\log N(u, \mathcal{T}, \hat{\rho}_S)} du \right).$$

Combining the above two inequalities yields

$$\left[ \mathbb{E} (\|S^o\|_{\mathcal{T}}^2) \right]^{1/2} \leq C \mathbb{E} \int_0^{\hat{D}_s} \sqrt{\log N(u, \mathcal{T}, \hat{\rho}_S)} du,$$

and  $\mathbb{E} \|S^o\|_{\mathcal{T}} \leq \left[ \mathbb{E} (\|S^o\|_{\mathcal{T}}^2) \right]^{1/2}$  which proves the theorem.  $\square$

The next theorem establishes the convergence of stochastic processes in the sense of Hoffmann-Jørgensen (1984). To prove that, we need to show finite

dimensional convergence to a Gaussian limit and stochastic equicontinuity. Assuming finite dimensional convergence to a Gaussian limit, it remains to show stochastic equicontinuity, in other words,

$$P^* \left( \sup_{\rho(s,t) \leq \epsilon} |S^o(s) - S^o(t)| > x \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

**THEOREM 6.2.** *Suppose that the summability of covariance condition (31) holds and the finite dimensional distribution of  $S^o$  has a Gaussian limit. If*

$$\sum_{i=1}^n \mathbb{E} \left( \|S_i(t)\|_{\mathcal{T}}^2 \right) = O(1), \quad (34)$$

$$\sum_{i=1}^n \mathbb{E} \left( I_{\{\|S_i(t)\|_{\mathcal{T}}^2 > u\}} \|S_i(t)\|_{\mathcal{T}}^2 \right) = o(1), \quad (35)$$

$$\int_0^{\epsilon(n)} \sqrt{\log N(u, \mathcal{T}, \widehat{\rho}_S)} du \xrightarrow{P^*} 0 \quad \text{whenever } \epsilon(n) \downarrow 0, \quad (36)$$

then  $\sum_{i=1}^n (S_i(t) - E[S_i(t)])$  is asymptotically equicontinuous and converges weakly to a Gaussian process.

**PROOF.** Let  $A_n \subset \mathbb{R}^n$  be a set of all vectors  $(S_1(t) - S_1(s), \dots, S_n(t) - S_n(s))$  for  $t, s$  such that  $\rho_S(t, s) \leq \epsilon(n)$ .

By the Markov's inequality, for every  $x > 0$ ,

$$P \left( \sup_{\rho(s,t) \leq \epsilon(n)} \left| \sum_{i=1}^n (S_i^o(s) - S_i^o(t)) \right| > x \right) \leq \frac{1}{x} \mathbb{E} \left( \sup_{\rho(s,t) \leq \epsilon(n)} \left| \sum_{i=1}^n (S_i^o(s) - S_i^o(t)) \right| \right)$$

Define  $F_n^2 = \sum_{i=1}^n \|S_i(t)\|_{\mathcal{T}}^2$ . Then by a similar argument in the proof of Pollard (1990) Theorem 10.6,

$$\mathbb{E} \left( \sup_{\rho(s,t) \leq \epsilon(n)} \left| \sum_{i=1}^n (S_i^o(s) - S_i^o(t)) \right| \right) \leq \mathbb{E} \left( |F_n| \Gamma(\delta_n / |2F_n|) \right)$$

where  $\Gamma(r^2) = \int_0^r \sqrt{\log N(u|F_n|, \mathcal{T}, \widehat{\rho}_S)} du$  and  $\delta_n = \sup_{A_n} \left| \sum_{i=1}^n a_i \right|$ .



Then for fixed  $r > 0$ , depending on  $|F_n| > r$  or not, the right hand side can be bounded by

$$\epsilon\Gamma(1)+\mathbb{E}\left(|F_n|\Gamma(\min\{1, \delta_n/2\epsilon\})\right) \leq \epsilon\Gamma(1)+\left[\left(\mathbb{E} F_n^2\right)\left(\mathbb{E} \Gamma^2(\min\{1, \delta_n/2\epsilon\})\right)\right]^{1/2}.$$

While the condition (34) ensures  $E|F_n|^2 < \infty$ , we need to show

$$\mathbb{E} \Gamma^2(\min\{1, \delta_n/2\epsilon\}) = o(1).$$

Since  $\Gamma$  is a continuous increasing function with  $\Gamma(0) = 0$ , our task is now to show that  $\delta_n$  converges to 0 in probability.

From here, we simply follow the classical empirical processes theory for the rest of the proof. See van der Vaart and Wellner (1996) Theorem 2.11.1 (See also Pollard 1990, Theorem 10.6) for the proof of convergence of  $\delta_n$  to 0 in probability.  $\square$

## References

- ALEXANDER, K.S. (1987). Central Limit theorems for stochastic processes under random entropy conditions. *Probab. Theory Rel. Fields.* **75**, 351–378.
- BERAN, R. (2000). REACT scatterplot smoothers : superefficiency through basis economy. *J. Amer. Statist. Assoc.* **63**, 155–171.
- BERAN, R., and DÜMBGEN, L. (1998). Modulation of Estimators and Confidence Sets. *Ann. Statist.*, **26**, 1826–1856.
- DUDLEY, R.M. (1987). Universal Donsker classes and metric entropy. *Ann. Prob.*, **15**, 1306–1326.
- EFROMOVICH, S. (1999). *Nonparametric Curve Estimation*. Springer, New York.
- GENOVESE, G. and WASSERMAN, L. (2002). Confidence Sets for Nonparametric Wavelet Regression. Technical Report. **768**, Carnegie Mellon University.

HOFFMANN-JØRGENSEN, J. (1984). Stochastic Processes on Polish Spaces. Unpublished.

MARRON, J.S. and WAND, M.P. (1992). *Exact Mean Integrated Squared Error*. *Ann. Statist.*, **20**, 712–736.

NUSSBAUM, M. (1996). *Asymptotic equivalence of density estimation and Gaussian white noise*. *Ann. Statist.*, **24**, 2399–2430.

POLLARD, D. (1990). *Empirical Processes : Theory and Applications*. NSF-CBMS Regional Conf. Ser. Prob. Statist. **2**. IMS, Hayward.

PORTNOY, S. (1986). On the Central Limit Theorem in  $R^p$  when  $p \rightarrow \infty$ . *Probab. Theory Rel. Fields*. **73**, 571–583.

SERFLING, R. J. (1980) *Approximation Theory of Mathematical Statistics*. Wiley, New York.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135–1151.

VAN DER VAART, A. W. and WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

W. Jang  
Institute of Statistics & Decision Sciences  
Duke University  
Durham, NC 27708  
wjang@stat.duke.edu

C. Genovese  
L. Wasserman  
Department of Statistics  
Carnegie Mellon University  
Pittsburgh, PA 15213