MINIMAX ESTIMATION WITH THRESHOLDING AND ASYMPTOTIC EQUIVALENCE FOR GAUSSIAN VARIANCE REGRESSION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Huibin Zhou

August 2004

MINIMAX ESTIMATION WITH THRESHOLDING AND ASYMPTOTIC EQUIVALENCE FOR GAUSSIAN VARIANCE REGRESSION

Huibin Zhou, Ph.D.

Cornell University 2004

Part I:

Many statistical practices involve choosing between a full model and reduced models where some coefficients are reduced to zero. Data were used to select a model with estimated coefficients. Is it possible to do so and still come up with an estimator always better than the traditional estimator based on the full model? The James–Stein estimator is such an estimator, having a property called minimaxity. However, the estimator considers only one reduced model, namely the origin. Hence it reduces no coefficient estimator to zero or every coefficient estimator to zero. In many applications including wavelet analysis, what should be more desirable is to reduce to zero only the estimators smaller than a threshold.

We construct such minimax estimators which perform thresholding. We apply our recommended estimator to the wavelet analysis and show that it performs the best among the well–known estimator aiming simultaneously at estimation and model selection. Some of our estimators are also shown to be asymptotically optimal.

Part II:

One of the most important statistical contributions of Lucien Le Cam is the asymptotic equivalence theory. A basic principle of asymptotic equivalence theory

is to approximate general statistical models by simple ones. A breakthrough of this theory was obtained by Nussbaum (1996) following the work of Brown and Low (1996). Nussbaum (1996) established the global asymptotic equivalence of the white-noise problem to the nonparametric density problem. The significance of asymptotic equivalence is that all asymptotically optimal statistical procedures can be carried over from one problem to the other when the loss function is bounded.

In this paper we established the asymptotic equvalence between the Gaussian variance regression problem and the Gaussian white noise problem under the Besov smoothness constraints. A multiresolution coupling methodology for the likelihood ratios (similar to the Hungarian construction) is used to establish asymptotic equivalence. For each resolution, our coupling approach is more elegant than the traditional quantile coupling methods; essentially we use quantile couplings between independent Beta's and independent normals. For the quantile coupling between a Beta random variable and a normal random variable, we establish a bound which improves the classical bound in KMT paper with a rate, which is of independent interest.

BIOGRAPHICAL SKETCH

Huibin Zhou was born on May 18, 1974 in Hubei, China. He entered the mathematics department studying mathematics education at Central China Normal University in 1991. After earning his B.S. degree in 1995, he attended Beijing University as a graduate student majoring in pure mathematics.

In 1998, Huibin Zhou came to Cornell to pursue his Ph.D degree in pure mathematics. In 1999, he switched to study statistics under the influence of professor J.T. Gene Hwang. Huibin Zhou earned his Ph.D degree from Cornell University in 2004.

Dedicated with great fondness to my mother and father who sacrificed everything for their children.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my advisors, Professor J.T. Gene Hwang and Professor Michael Nussbaum, for being nearly perfect advisors. Their friendship, guidance, and encouragement have given me the greatest experience.

Thanks to Professor Robert Strichartz and Professor Marty Wells for being on my Ph.D. committee. Their ideas and advice have been a valuable contribution to this work.

Thanks to Professors Lawrence D. Brown, Tony T. Cai, Eugene B. Dynkin and Mark G. Low for their support and tremendous help during my years of Ph.D study.

I am grateful and indebted to Lorin DeYoung, Jianjin Gui, Brian Hwang, Jing-huei Hwang, Jing Qiu, Marlene Sheridan and Gongfu Zhou for their friendship and support.

I would also like to thank my mother, father, and brothers. Their love and support have carried me to where I am today. In particular, my eldest brother sacrificed much for me and my whole family.

I want to thank my wife, Bo Bi, for her love, support and understanding which brought to my life unimagined happiness. I want to thank my sons, Michael and Jeffrey, who give me endless joy.

TABLE OF CONTENTS

1 Minimax Estimation with Thresholding and Its Application to Wavelet Analysis **1**

1.1 Introduction 1

1.2 New Estimators for a Canonical Model 2

1.3 Approximate Bayesian Justification. 11

1.4 Empirical Bayes Justification. 12

1.5 Connection to the Wavelet Analysis and the Numerical Results. . . . 14

1.6 Asymptotic Optimality. 17

1.7 Appendix. 18

Bibliography **35**

2 Asymptotic Equivalence Theory of the Gaussian Variance Regression Experiment **37**

2.1 Introduction 37

2.2 The Structure of Likelihood 43

2.3 Construction of the Likelihoods on the Same Probability Space . . . 50

2.4 Proof of the Main Result 54

2.4.1 An Upper Bound of N_{21} 55

2.4.2 An upper bound of N_{22} 62

2.5 The Preliminary Estimator 69

2.6 Quantile Coupling for Beta Distributions 73

2.7 Appendix 82

Bibliography **95**

LIST OF FIGURES

1.1 Relative Frequentist risk of the proposed estimator (1.12) to p=50. 29
1.2 The curves represent the true curves f(t)in (1.18). 30
1.3 In each of the six casescorresponding to Blocks, Bumps, etc., the eight curves plot the risk function, from top to the bottom, when $n = 64, 128, \ldots, 8192$. For each curve (see for example, the top curve on the left), the circles "o" from left to the right give, with respect to n, the relative risks of VisuShrink, Block James–Stein, SureShrink, and the proposed methods (1.12) and (1.20). 31
1.4 Solid lines represent the true curves, wheredotted lines represent the curves corresponding to various estimators. The simulated risk is based on 500 simulations. . . 32
1.5 Proposed Estimator (1.12) Applied to Reconstruct Figure 1.2. 33
1.6 James-Stein Positive Part Applied to Reconstruct Figure 1.2. 34

Chapter 1

Minimax Estimation with Thresholding and Its Application to Wavelet Analysis

1.1 Introduction

In virtually all statistical activities, one constructs a model to summarize the data. Not only could the model provide a good and effective way of summarizing the data, the model if correct often provides more accurate prediction. This point has been argued forcefully in Gauch (1993). Is there a way to use the data to select a reduced model so that if the reduced model is correct the model based estimator will improve on the naive estimator (constructed using a full model) and yet never do worse than the naive estimator even if the full model is actually the only correct model? James–Stein estimation (1961) provide such a striking result under the normality assumption. Any estimator such as the James-Stein estimator that does no worse than the naive estimator is said to be minimax. See the precise discussion right before Lemma 1.1 of Section 1.2. The problem with the James–Stein positive part estimator is however that it selects only between two models: the origin and the full model. It is possible to construct estimators similar to James–Stein positive part to select between the full model and another linear subspace. However it always chooses between the two. The nice idea of George (1986a,b) in multiple shrinkage does allow the data to choose among several models; it however does not do thresholding as is the aim of the paper.

Models based on wavelets are very important in many statistical applications. Using these models involves model selection among the full model or the models with smaller dimensions where some of the wavelet coefficients are zero. Is there a way to select a reduced model so that the estimator based on it does no worse in any case than the naive estimator based on the full model, but improves substantially upon the naive estimator when the reduced model is correct? Again, the James–Stein estimator provides such a solution. However it selects either the origin or the full model. Furthermore, the ideal estimator should do thresholding, namely it gives zero as an estimate for the components which are smaller than a threshold, and preserves (or shrinks) the other components. However, to the best knowledge of the authors, no such minimax estimators have been constructed. In this paper, we provide minimax estimators which perform thresholding simultaneously.

Section 1.2 develops the new estimator for the canonical form of the model by solving Stein's differential inequality. Sections 1.3 and 1.4 provide an approximate Bayesian justification and an empirical Bayes interpretation. Section 1.5 applies the result to the wavelet analysis. The proposed method outperforms several prominent procedures in the statistical wavelet literature. Asymptotic optimality of some of our estimators is established in Section 1.6.

1.2 New Estimators for a Canonical Model

In this section, we shall consider the canonical form of the problem of a multinormal mean estimation problem under the squared error loss. Hence we shall assume that our observation

$$Z = (Z_1, \ldots, Z_d) \sim N(\theta, I)$$

is a d–dimensional vector consisting of normal random variable with mean $\theta = (\theta_1, \ldots, \theta_d)$, and a known covariance identity matrix I. The case when the variance of Z_i is not known will be discussed briefly at the end of Section 1.5.

The connection of this problem with wavelet analysis will be pointed out in Sections 1.5 and 1.6. In short Z_i and θ_i represent the wavelet coefficients of the data and the true curve in the same resolution, respectively. Furthermore d is the dimension of a resolution. For now, we shall seek an estimator of θ based on Z. We shall, without loss of generality, consider an estimator of the form $\delta(Z) = (\delta_1(Z), \ldots, \delta_d(Z))$, where

$$\delta_i(Z) = Z_i + g_i(Z)$$

where $g(Z) : R^d \to R$ and search for $g(Z) = (g_1(Z), \ldots, g_d(Z))$. To insure that the new estimator (perhaps with some thresholding) does better than Z (which does no thresholding), we shall compare the *risk* of $\delta(Z)$ to the risk of Z with respect to the ℓ_2 norm. Namely

$$E\|\delta(Z) - \theta\|^2 = E\sum_{i=1}^{d}(\delta_i(Z) - \theta_i)^2.$$

It is obvious that the risk of Z is then d. We shall say one is *as good as* the other if the former has a risk no greater than the latter for every θ. Moreover, one *dominates* the other if it is as good as the other and has smaller risk for some θ. Also we shall say that an estimator *strictly dominates* the other if the former has a smaller risk for every θ. Note that Z is a minimax estimator, i.e., it minimizes $\sup_\theta E|\delta^0(Z) - \theta|^2$ among all $\delta^0(Z)$. Consequently any $\delta(Z)$ is as good as Z if and only if it is minimax.

To construct an estimator that dominates Z, we use the following lemma.

Lemma 1.1 *(Stein 1981) Suppose that $g : R^d \to R^d$ is a measurable function with $g_i(\cdot)$ as the ith component. If for every i, $g_i(\cdot)$ is almost differentiable with respect* *to ith component and*

$$E\left(\left|\frac{\partial}{\partial Z_i}g_i(Z)\right|\right) < \infty, \ \textit{for } i = 1, \ldots, d,$$

then

$$E_\theta\|Z + g(Z) - \theta\|^2 = E_\theta\{d + 2\nabla \cdot g(Z) + \|g(Z)\|^2\},$$

where $\nabla \cdot g(Z) = \sum_{i=1}^d \frac{\partial g_i(Z)}{\partial Z_i}$. *Hence if* $g(Z)$ *solves the differential inequality*

$$2\nabla \cdot g(Z) + \|g(Z)\|^2 < 0, \tag{1.1}$$

the estimator $Z + g(Z)$ *strictly dominates* Z*.*

Remark 1.1 *:* $g_i(z)$ *is said to be almost differentiable with respect to* z_i*, if for almost all* z_j*,* $j \neq i$*,* $g_i(z)$ *can be written as a one dimensional integral of a function with respect to* z_i*. For such* z_j*'s,* $j \neq i$*,* $g_i(Z)$ *is also called absolutely continuous with respect to* z_i *in Berger (1980).*

To motivate the proposed estimator, note that the James–Stein positive estimator has the form

$$\widehat{\theta}_i^{JS} = \left(1 - \frac{d-2}{\|Z\|^2}\right)_+ Z_i$$

when $c_+ = \max(c, 0)$ for any number c. This estimator, however, truncates independently of the magnitude of $|Z_i|$. Indeed, it truncates all or none of the coordinates. To construct an estimator that truncates only the coordinate with small $|Z_i|$'s, it seems necessary to replace $d-2$ by a decreasing function $h(|Z_i|)$ of $|Z_i|$ and consider

$$\widehat{\theta}_i^{+} = \left(1 - \frac{h(|Z_i|)}{D}\right)_+ Z_i$$

where D, independently of i, is yet to be determined. (In a somewhat different approach, Beran and Dümbgen (1998) constructs a modulation estimator corresponding to a monotonic shrinkage factor.) With such a form, $\widehat{\theta}_i^+ = 0$ if $h(|Z_i|) \geq D$, which has a better chance of being satisfied when $|Z_i|$ is small.

We consider a simple choice $h(|Z_i|) = a|Z_i|^{-2/3}$, and let $D = \Sigma|Z_i|^{4/3}$. This leads to the untruncated version $\widehat{\theta}$ with the ith component

$$\widehat{\theta}_i(Z) = Z_i + g_i(Z) \text{ where } g_i(Z) = -aD^{-1} sign(Z_i)|Z_i|^{1/3}. \tag{1.2}$$

Here and later $sign(Z_i)$ denotes the sign of Z_i. It is possible to use other decreasing functions $h(|Z_i|)$ and other D.

In general, we consider, for a fixed $\beta \leq 2$, an estimator of the form

$$\widehat{\theta}_i = Z_i + g_i(Z), \tag{1.3}$$

where

$$g_i(Z) = -a\frac{sign(Z_i)|Z_i|^{\beta-1}}{D} \quad \text{and} \quad D = \sum_{i=1}^{d} |Z_i|^{\beta}. \tag{1.4}$$

Although at first glance, it may seem hard to justify this estimator, it has a Bayesian and Empirical Bayes justification in Sections 1.3 and 1.4. It contains, as a special case with $\beta = 2$, the James-Stein estimator. Now we have

Theorem 1.1 *For $d \geq 3$ and $1 < \beta \leq 2$, $\widehat{\theta}(Z)$ is minimax if and only if*

$$0 < a \leq 2(\beta - 1) \inf_{\theta} \frac{E_\theta\left(D^{-1}\sum_{i=1}^{d}|Z_i|^{\beta-2}\right)}{E_\theta(D^{-2}\sum_{i=1}^{d}|Z_i|^{(2\beta-2)})} - 2\beta.$$

Proof: : Obviously for $Z_j \neq 0, \forall\, j \neq i$, $g_i(Z)$ can be writen as the one–dimensional integral of

$$\frac{\partial}{\partial Z_i} g_i(Z) = \beta(-a)(-1)D^{-2}|Z_i|^{(2\beta-2)} + (\beta-1)(-a)D^{-1}(|Z_i|^{\beta-2})$$

with respect to Z_i. (The only concern is at $Z_i = 0$.) Consider only nonzero Z_j's, $j \neq i$. Since $\beta > 1$, this function however is integrable with respect to Z_i even over an integral including zero. It takes some effort to prove that $E(|\frac{\partial}{\partial Z_i} g_i(Z)|) < \infty$. However one only needs to focus on Z_j close to zero. Using the spherical–like transformation $r^2 = \sum |Z_i|^\beta$, we may show that if $d \geq 3$ and $\beta > 1$ both terms in the above displayed expression are integrable.

Now

$$\|g(Z)\|^2 = a^2 D^{-2} \sum_{i=1}^{d} |Z_i|^{2\beta-2}.$$

Hence

$$E_\theta \|Z + g(Z) - \theta\|^2 \leq d, \text{ for every } \theta,$$

if and only if,

$$E_\theta \{2\nabla \cdot g(Z) + \|g(Z)\|^2\} \leq 0, \text{ for every } \theta,$$

i.e.,

$$E_\theta \left(\begin{array}{c} a\left((2\beta)D^{-2}\sum_{i=1}^{d}|Z_i|^{(2\beta-2)} - (2\beta-2)D^{-1}\sum_{i=1}^{d}|Z_i|^{\beta-2}\right) \\ +a^2 D^{-2}\sum_{i=1}^{d}|Z_i|^{2\beta-2} \end{array} \right)$$
$$\leq 0, \qquad \text{for every } \theta, \tag{1.5}$$

which is equivalent to the condition stated in the Theorem.

Theorem 1.2 *The estimator $\widehat{\theta}(Z)$ with the ith component given in (1.2) and (1.3) is minimax provided $0 < a \leq 2(\beta-1)d - 2\beta$ and $1 < \beta \leq 2$. Unless $\beta = 2$ and a is taken to the upper bound, otherwise $\widehat{\theta}(Z)$ dominnates Z.*

Proof: : By the correlation inequality

$$d\sum_{i=1}^{d}|Z_i|^{2\beta-2} \leq \left(\sum_{i=1}^{d}|Z_i|^{(\beta-2)}\right)\left(\sum_{i=1}^{d}|Z_i|^{\beta}\right).$$

Strict inequality holds almost surely if $\beta < 2$. Hence

$$\frac{E_\theta\left(D^{-1}\sum_{i=1}^{d}|Z_i|^{\beta-2}\right)}{E_\theta(D^{-2}\sum_{i=1}^{d}|Z_i|^{2\beta-2})} \geq \frac{E_\theta D^{-1}\sum|Z_j|^{\beta-2}}{\frac{1}{d}E_\theta D^{-1}\sum|Z_i|^{\beta-2}} = d.$$

Hence if $0 < a \leq 2(\beta - 1)d - 2\beta$, then the condition in Theorem 2 is satisfied, implying minimaxity of $\widehat{\theta}(Z)$. The rest of the statement of the theorem is now obvious.

The following theorem is a generalization of Theorem 6.2 on page 302 of Lehmann (1983) and Theorem 5.4 on page 356 of Lehmann and Casella (1998). It shows that taking the positive part will improve the estimator componentwise. Specifically for an estimator $(\widetilde{\theta}_1(Z), \ldots, \widetilde{\theta}_d(Z))$ where

$$\widetilde{\theta}_i(Z) = (1 - h_i(Z))Z_i,$$

the positive part estimator of $\widetilde{\theta}_i(Z)$ is denoted as

$$\widetilde{\theta}_i^+(Z) = (1 - h_i(Z))_+ Z_i.$$

Theorem 1.3 *Assume that $h_i(Z)$ is symmetric with respect to the ith coordinate, then*

$$E_\theta(\theta_i - \widetilde{\theta}_i^+)^2 \leq E_\theta(\theta_i - \widetilde{\theta}_i)^2.$$

Furthermore, if

$$P_\theta(h_i(Z) > 1) > 0, \tag{1.6}$$

then

$$E_\theta(\theta_i - \widetilde{\theta}_i^+)^2 < E_\theta(\theta_i - \widetilde{\theta}_i)^2.$$

Proof: : Simple calculation shows that

$$E_\theta(\theta_i - \widetilde{\theta}_i^+)^2 - E_\theta(\theta_i - \widetilde{\theta}_i)^2 = E_\theta((\widetilde{\theta}_i^+)^2 - \widetilde{\theta}_i^2) - 2\theta_i E_\theta(\widetilde{\theta}_i^+ - \widetilde{\theta}_i). \tag{1.7}$$

Let's calculate the expectation by conditioning on $h_i(Z)$. For $h_i(Z) \leq 1$, $\widetilde{\theta}_i^+ = \widetilde{\theta}_i$. Hence it is sufficient to condition on $h_i(z) = b$ where $b > 1$ and show that

$$E_\theta((\widetilde{\theta}_i^+)^2 - \widetilde{\theta}_i^2 \mid h_i(Z) = b) - 2\theta_i E_\theta(\widetilde{\theta}_i^+ - \widetilde{\theta}_i \mid h_i(Z) = b) \leq 0,$$

or equivalently,

$$-E_\theta(\widetilde{\theta}_i^2 \mid h_i(Z) = b) + 2\theta_i(1-b)E_\theta(Z_i \mid h_i(Z) = b) \leq 0.$$

Obviously, the last inequality is satisfied if we can show

$$\theta_i E_\theta(Z_i \mid h_i(Z) = b) \geq 0.$$

We may further condition on $Z_j = z_j$ for $j \neq i$ and it suffices to establish

$$\theta_i E_\theta(Z_i \mid h_i(Z) = b, Z_j = z_j, j \neq i) \geq 0. \tag{1.8}$$

Given that $Z_i = z_j$, $j \neq i$, consider only the case where $h_i(Z) = b$ has solutions. Due to symmetry of $h_i(Z)$, these solutions are in pairs. Let $\pm y_k$, $k \in K$, denote the solutions. Hence the left hand side of (1.8) equals

$$\begin{aligned} \theta_i E_\theta(Z_i \quad &| \quad Z_i = \pm y_k, k \in K) \\ &= \sum_{k \in K} \theta_i E_\theta(Z_i \mid Z_i = \pm y_k) P_\theta(Z_i = \pm y_k \mid Z_i = \pm y_k, k \in K). \end{aligned}$$

Note that

$$\theta_i E_\theta(Z_i \mid Z_i = \pm y_k) = \frac{\theta_i y_k e^{y_k \theta_i} - \theta_i y_k e^{-y_k \theta_i}}{e^{y_k \theta_i} + e^{-y_k \theta_i}}, \tag{1.9}$$

which is symmetric in $\theta_i y_k$ and is increasing for $\theta_i y_k > 0$. Hence (9) is bounded below by zero, a bound obtained by substituting $\theta_i y_k = 0$ in (1.9). Consequently we establish that (1.7) is nonpositive, implying the domination of $\widetilde{\theta}_i^+$ over $\widetilde{\theta}_i$.

The strict inequality of the theorem can be established by noting that the right hand side of (1.7) is bounded above by $E_\theta[(\widetilde{\theta}_i^+)^2 - \widetilde{\theta}_i^2]$ which by (1.6) is strictly negative.

Theorem 1.3 implies the following Corollary.

Corollary 1.1 *Under the assumption of Theorem 3, $\widehat{\theta}^+$ with ith component*

$$\widehat{\theta}^+ = (1 - aD^{-1}|Z_i|^{\beta-2})_+ Z_i \tag{1.10}$$

strictly dominates Z.

It is interesting to note that estimator (1.10), for $\beta < 2$, does give zero as the estimator when $|Z_i|$ are small. When applied to the wavelet analysis, it truncates the small wavelet coefficients and shrinks the large wavelet coefficients. The estimator lies in a data chosen reduced model.

Moreover, for $\beta = 2$, Theorem 1.2 reduces to the classical result of Stein (1981) and (1.10) to the positive part James-Stein estimator. The upper bound of a for domination stated in Theorem 3 works only if $\beta > 1$ and $d > \beta/(\beta - 1)$. We know that for $\beta \leq \frac{1}{2}$, $\widehat{\theta}$ fails to dominate Z because of the calculations leading to (1.11) below. We are unable to prove that $\widehat{\theta}$ dominates Z for $\frac{1}{2} < \beta \leq 1$. However, for such β's, $\widehat{\theta}$ has a smaller Bayes risk than Z if the condition (1.11) below is satisfied.

A Remark about an Explicit Formula for a:

In wavelet analysis, a vast majority of the wavelet coefficients of a reasonably smooth function are zero. Consequently, it seems good to choose an estimator that shrinks a lot and hence using a larger than the upper bound in Theorem 3 is desirable. Although Theorem 2 provides the largest possible a for domination in the frequentist sense, the bound is difficult to evaluate in computation and hence difficult to use in a real application. Hence we took an alternative approach by assuming that θ_i are i.i.d. $N(0, \tau^2)$. Note that the difference of the Bayes risk of $\widehat{\theta}$ and Z equals $E(\mathcal{D})$, where

$$\begin{aligned} \mathcal{D} &= \sum_{i=1}^{p} \left((Z_i + g_i(Z) - \theta_i)^2 - (Z_i - \theta_i)^2 \right) \\ &= \sum_{i=1}^{p} \left(2\,(Z_i - \theta_i)\, g_i(Z) + g_i^2(Z) \right) \end{aligned}$$

To calculate the expectation with respect to Z_i and θ_i, we first calculate the conditional expectation given Z_i. Since $E(\theta_i|Z_i) = \frac{\tau^2 Z_i}{1+\tau^2}$, we obtain

$$\begin{aligned} E(\mathcal{D}) &= E[E(\mathcal{D}|Z_1,\ldots,Z_p)] = E\left(\sum_{i=1}^{p}\left(2\frac{Z_i}{1+\tau^2}g_i(Z)+g_i^2(Z)\right)\right) \\ &= E\left(\frac{a^2}{D^2}\sum_{i=1}^{p}|Z_i|^{2\beta-2} - \frac{2a}{1+\tau^2}\right). \end{aligned}$$

Note that $E(\mathcal{D}) \le 0$, if

$$0 \le a \le \frac{2}{1+\tau^2}E\left(\frac{\sum_{i=1}^{p}|Z_i|^{2\beta-2}}{D^2}\right)$$

where the expectation is taken over Z_i which are i.i.d. and

$$Z_i \sim N\left(0, 1+\tau^2\right).$$

Let $\xi_i = Z_i/\sqrt{1+\tau^2}$ and consequently $\xi_i \sim N_p(0,1)$. Thus the estimator (1.2) and (1.3) has a smaller Bayes risk than Z for all τ^2 if and only if

$$0 < a < a_\beta = 2/E\left[\sum_{i=1}^{d}|\xi_i|^{2\beta-2}/\left(\sum_{i=1}^{d}|\xi_i|^{\beta}\right)^2\right] \tag{1.11}$$

where ξ_i are i.i.d. standard normal random variables.

What is the value of a_β? It is easy to numerically calculate the bound a_β by simulating ξ_i, which we did for a up to 100. It is shown that a_β, $\beta = \frac{4}{3}$ is at least as big as $(5/3)(d-2)$. Using Berger's (1976) tail minimaxity argument, we come to the conclusion that $\widehat{\theta}^+$, with the ith component

$$\widehat{\theta}_i = \left(1 - \frac{5/3(d-2)Z_i^{-2/3}}{\sum_{i=1}^{d} Z_i^{4/3}}\right)_+ Z_i \tag{1.12}$$

would possibly dominate Z. For various d's including $d = 50$, this was shown to be true numerically.

To derive a general formula for a_β for all β, we then establish that the limit of a_β/d as $d \to \infty$ equals, for $1/2 < \beta < 2$,

$$C_\beta = 4[\Gamma((\beta+1)/2)]^2/[\sqrt{\pi}\Gamma((2\beta-1)/2)]. \tag{1.13}$$

It may be tempting to use $(d-2)C_\beta$. However we recommend

$$a = (0.97)(d-2)C_\beta, \tag{1.14}$$

so that at $\beta = 4/3$, (1.14) becomes $(5/3)(d-2)$. Berger's tail minimaxity argument and many numerical studies indicate that this a enables (1.10) to have a better risk than Z. For $d = 50$, it is shown that $\widehat{\theta}^+$ dominates Z in Figure 1.1.

1.3 Approximate Bayesian Justification.

It would seem interesting to justify the proposed estimation from a Bayesian's point of view. To do so, we consider a prior of the form

$$\begin{aligned} \pi(\theta) &= 1,\ \|\theta\|_\beta \le 1 \\ &= 1/(\|\theta\|_\beta)^{\beta c},\ \|\theta\|_\beta > 1 \end{aligned}$$

where $\|\theta\|_\beta = (\sum \|\theta_i\|^\beta)^{1/\beta}$, and c is a positive constant which can be specified to match the constant a in (1.10). In general the Bayes estimator is given by

$$Z + \nabla \log m(Z)$$

where $m(Z)$ is the marginal probability density function of Z. Namely,

$$m(Z) = \int \cdots \int \frac{e^{-\frac{1}{2}\|Z-\theta\|^2}}{(\sqrt{2\pi})^d} \pi(\theta) d\theta.$$

The following approximation follows from Brown (1971), which asserts that $\nabla \log m(Z)$ can be approximated by $\nabla \log \pi(Z)$. The proof is given in the Appendix.

Theorem 1.4 *With $\pi(\theta)$ and $m(X)$ given above,*

$$\lim_{|Z_i| \to +\infty} \frac{\nabla_i \log m(Z)}{\nabla_i \log \pi(Z)} = 1.$$

Hence by Theorem 6, the ith component of the Bayes estimator equals approximately

$$Z_i + \nabla_i \log \pi(Z) = Z_i - \frac{c\beta |Z_i|^{\beta-1} sign(Z_i)}{\sum |Z_i|^\beta}.$$

This is similar to the untruncated version of $\widehat{\theta}$ in (1.2) and (1.3).

1.4 Empirical Bayes Justification.

Based on several signals and images, Mallat (1989) proposed a prior for the wavelelet coefficients θ_i as the exponential power distribution with the probability density function (p.d.f.) of the form

$$f(\theta_i) = ke^{-|\frac{\theta_i}{\alpha}|^\beta} \tag{1.15}$$

where α and $\beta < 2$ are positive constants and

$$k = \beta/(2\alpha\Gamma(1/\beta))$$

is the normalization constant. See also Vidakovic (1999, p.194). Using method of moments, Mallat estimated value of α and β to be 1.39 and 1.14 for a particular graph. However, α and β are typical unknown.

It seems reasonable to derive an Empirical Bayes estimator based on this class of prior distributions. First we assume that α is known. Then the Bayes estimator of θ_i is

$$Z_i + \frac{\partial}{\partial Z_i} \log m(Z).$$

Similar to the argument in Theorem 1.4 and noting that for $\beta < 2$,

$$e^{-|\theta_i + Z_i|^\beta/\alpha^\beta} / e^{-|\theta_i|^\beta/\alpha^\beta} \to 1 \text{ as } \theta_i \to \infty,$$

the Bayes estimator can be approximated by

$$Z_i + \frac{\partial}{\partial Z_i} \log \pi(Z_i) = Z_i - \frac{\beta}{\alpha^\beta} |Z_i|^{\beta-1} sign(Z_i). \tag{1.16}$$

Note that, under the assumption that α is known, the above expression is also the asymptotic expression of the maximum likelihood estimator of θ_i by maximizing the joint p.d.f. of (Z_i, θ_i). See Proposition 1 of Antoniadis, Leporini and Desquet (2002) as well as (8.23) of Vidakovic (1999). In the latter reference, the sign of Z_i of (1.16) is missing due to a minor typographic error.

Since α is unknown, it seems reasonable to replace α in (1.16) by an estimator. Assume that θ_i's are observable. Then by (1.15) the joint density of $(\theta_1, \ldots, \theta_d)$ is

$$\left[\left(\frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \right)^d e^{-\Sigma\left(\frac{|\theta_i|^\beta}{\alpha^\beta}\right)} \right].$$

Differentiating this p.d.f. with respect to α gives the maximum likelihood estimator of α^β as

$$(\beta \Sigma |\theta_i|^\beta)/d. \tag{1.17}$$

However since θ_i is unknown and hence the above expression can be further estimated by (1.16). For $\beta < 2$, the second term in (1.16) has a smaller order than the first when $|Z_i|$ is large. Replacing θ_i by the dominating first term Z_i in (1.16) leads to an estimator of α^β as $(\beta\Sigma|Z_i|^\beta)/d$.

Substituting this into (16) gives

$$Z_i - \frac{d}{\Sigma|Z_i|^\beta} |Z_i|^{\beta-1} sign(Z_i)$$

which is exactly estimator $\widehat{\theta}_i$ in (1.2) and (1.3) with $a = d$. Hence we have succeeded in deriving $\widehat{\theta}_i$ as an Empirical Bayes estimator when Z_i is large.

1.5 Connection to the Wavelet Analysis and the Numerical Results.

Wavelets have become a very important tool in many areas including Mathematics, Applied Mathematics, Statistics, and signal processing. It is also applied to numerous other areas of science such as chemometrics and genetics.

In statistics, wavelets have been applied to function estimation with amazing results of being able to catch the sharp change of a function. Celebrated contributions by Donoho and Johnstone (1994 and 1995) focus on developing thresholding techniques and asymptotic theories. In the 1994 paper, relative to the oracle risk, their VisuShrink was shown to be asymptotically optimal. Further in 1995's paper, the expected squared error loss of their SureShrink is shown to achieve the global asymptotic minimax rate over Besov spaces. Cai (1999) improved on their result by establishing that the Block James–Stein (BlockJS) thresholding achieve exactly the asymptotic global or local minimax rate over various classes of Besov spaces.

Now specifically let $Y = (Y_1, \ldots, Y_n)'$ be samples of a function f, satisfying

$$Y_i = f(t_i) + \varepsilon_i \tag{1.18}$$

where $t_i = (i-1)/n$ and ε_i are independently identically distributed (i.i.d.) $N(0, \sigma^2)$. Here σ^2 is assumed to be known and is taken to be one without loss of generality. See a comment at the end of the paper regarding the unknown σ case. One wishes to choose an estimate $\widehat{f} = (\widehat{f}(t_1), \ldots, \widehat{f}(t_n))$ so that its risk function

$$E\|\widehat{f} - f\|^2 = E\sum_{i=1}^{n}(\widehat{f}(t_i) - f(t_i))^2, \tag{1.19}$$

is as small as possible. Many discrete wavelet transformations are orthogonal transformations. See Donoho and Johnstone (1995). Consequently, there exists an orthogonal matrix W, such that the wavelet coefficients of Y and f are $Z = WY$ and $\theta = Wf$. Obviously the components Z_i of Z are independent, having a normal distribution with mean θ_i and standard deviation 1. Hence previous sections apply and exhibit many good estimators of θ. Note that, by orthogonality of W, for any estimator $\delta(Z)$ of θ, its risk function is identical to $W'\delta(Z)$ as an estimator of $f = W'\theta$. Hence the good estimators in previous sections can be inversely transformed to estimate f well.

In all the applications to wavelets discussed in this paper, the estimators (including our proposed estimator) apply separately to the wavelet coefficients of the same resolution. Hence in (1.12), for example, d is taken to be the number of coefficients of a resolution when applied to the resolution. In all the literature that we are aware of, this has been the case as well.

In addition to considering the estimator (1.12), which is a special case of (1.10) with $\beta = 4/3$, we also propose a modification (1.10) with an estimated β. The estimator $\widehat{\beta}$ for β is constructed by minimizing, for each resolution, the Stein's unbiased risk estimator (SURE) for the risk of (1.10). The quantity SURE is basically the expression inside the expectation on the right hand side of (1.27) summing over i, $1 \le i \le d$, except that a is replaced by a_β. (Note that D in (1.27) depends on β as well.) The resultant estimator is denoted as

$$\widehat{\theta}^S = (1.10) \text{ with } \beta \text{ replaced by } \widehat{\beta}. \tag{1.20}$$

Figure 1.2 gives six true curves (made famous by Donoho and Johnstone) from which the data are generated. For these six cases, Figure 1.3 plots the ratios of the risks of the aforementioned estimators to n, the risk of Y. Since most relative risks are less than one, this indicates that most estimators perform better than the raw data Y. Our estimators $\widehat{\theta}^+$ in (1.12) and $\widehat{\theta}^S$ in (20), however, are the ones that are consistently better than Y. Furthermore, our estimators $\widehat{\theta}^+$ and $\widehat{\theta}^S$ virtually dominate all the other estimators in risk. Generally, $\widehat{\theta}^S$ performs better than $\widehat{\theta}^+$ virtually in all cases.

As shown in Figure 1.3, the difference in risks between $\widehat{\theta}^+$ and $\widehat{\theta}^S$ are quite minor. Since $\widehat{\theta}^+$ is computationally less intensive, we focus on $\widehat{\theta}^+$ for the rest of the numerical studies.

Picturewise, our estimator does slightly better than other estimators. See Figure 1.4 for an example. Note that the picture corresponding to $\widehat{\theta}^+$ distinguishes most clearly the first and second bumps from the right.

Based on asymptotic calculation, the next section also recommends a choice of a in (1.21). It would seem interesting to comment on its numerical performance. The difference between the a's defined in (1.14) and (1.22) are very small when $64 \leq n \leq 8192$ and when β is estimated by minimizing SURE. Consequently, for such β, the risk functions of the two estimators with different a's are very similar, with a difference virtually bounded by 0.02. The finite sample estimator (where a is defined in (1.14)) has a smaller risk about 75% of the times.

James–Stein estimator produces very attractive risk functions, sometimes as good as the proposed estimator (1.12). However, it does not seem to produce good graphs. Compare Figures 1.5 and 1.6.

In the simulation studies, we use the procedures MultiVisu and MultiHybrid which are VisuShrink and SureShrink in WaveLab802. See http://playfair.stanford.edu/∼wavelab. We use Symmlet 8 to do wavelet transformation. In Figure 1.3, signal to noise ratio (SNR) is taken to be 3. Results are similar for other SNR's. To include block thresholding result of Cai (1999), we choose the lowest integer resolution level $j \geq \log_2(\log n) + 1$.

A comment about the case where σ^2 is not known to be one.

When σ is known and is not equal to one, a simple transformation applied to the problem suggest that (1.10) be modified with a replaced by $a\sigma^2$. When σ is unknown, one could then estimate σ by $\widehat{\sigma}$, the proposed estimator for σ in Donoho and Johnston (1995, page 1218). With this modification in (1.12) or with a SURE estimated β, the resultant estimators are not minimax according to some numerical simulations. However, they still perform the best or nearly the best among all the estimators studied in Figure 1.3.

1.6 Asymptotic Optimality.

To study the asymptotic rate of a wavelet analysis estimator, it is customary to assume the model

$$Y_i = f(t_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{1.21}$$

where $t_i = (i-1)/n$ and ε_i are assumed to be i.i.d. $N(0,1)$. The estimator $\widehat{f}$ for $f(\cdot)$ that can be proved asymptotically optimal applies estimator (10) with

$$a = d(2\ln d)^{(2-\beta)/2} m_\beta, \ 0 \le \beta \le 2, \tag{1.22}$$

and

$$m_\beta = E|\varepsilon_i|^\beta = 2^{\beta/2}\Gamma((\beta+2)/2)\sqrt{\pi},$$

to the wavelet coefficients Z_i of each resolution with dimensionality d of the wavelet transformation of Y_i's. After applying the estimator to each resolution one at a time to come up with the new wavelet coefficient estimators, one then uses the wavelet base function to obtain one function $\widehat{f}$ in the usual way.

To state the theorem, we use $B^\alpha_{p,q}$ to denote the Besov space with smoothness α and shape parameters p and q. The definition of the Besov class $B^\alpha_{p,q}(M)$ with respect to the wavelet coefficients are given in (1.42). Now the asymptotic theorem is given below.

Theorem 1.5 *Assume that the wavelet ψ is γ–regular, i.e., ψ has γ vanishing moments and γ continuous derivatives. Then there exists a constant C independent of n and f such that*

$$\sup_{f \in B^{\alpha}_{p,q}(M)} E \int_0^1 |f(t) - \widehat{f}(t)|^2 dt \le C(\ln n)^{1-\beta/2} n^{-2\alpha/(2\alpha+1)}, \qquad (1.23)$$

for all $M > 0$, $0 < \alpha < r$, $q \ge 1$ and $p > \max(\beta, \frac{1}{\alpha}, 1)$.

The asymptotic optimality stated in (1.23) is as good as what has been established for hard and soft thresholding estimators in Donoho and Johnstone (1994), the Garrott method in Gao (1998) and Theorem 4 in Cai (1999) and SCAD method in Antoniadis and Fan (2001). However, the real advantage of our estimator is in the finite sample risk as reported in Section 1.5. Also our estimators are constructed to be minimax and hence have finite risk functions uniformly smaller than the risk of Z. This estimator $\widehat{\theta}^A$ for $\beta = 4/3$ however has a risk very similar to (1.12). See Section 1.5.

1.7 Appendix.

Proof of Theorem 1.4. Assume that $|Z_i| > 1$. We have

$$\lim_{|Z_i| \to \infty} \frac{\nabla_i \log m(Z)}{\nabla_i \log \pi(Z)} = \lim_{|Z_i| \to +\infty} \frac{\pi(Z)}{m(Z)} \cdot \frac{\frac{\partial}{\partial Z_i} m(Z)}{\frac{\partial}{\partial Z_i} \pi(Z)} .$$

We shall prove only

$$\lim_{|Z_i| \to \infty} \frac{m(Z)}{\pi(Z)} = 1,$$

since

$$\lim_{|Z_i|\to\infty} \frac{\frac{\partial}{\partial Z_i} m(Z)}{\frac{\partial}{\partial Z_i}\pi(Z)} = 1$$

can be similarly established.

Now

$$\begin{aligned} m(Z) &= \int\cdots\int \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\|Z-\theta\|^2}\pi(\theta)d\theta \\ &= \int\cdots\int_{\|\theta\|_\beta\le 1} \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\|Z-\theta\|^2} d\theta \\ &\quad + \int\cdots\int_{\|\theta\|_\beta > 1} \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\|Z-\theta\|^2} \frac{1}{\|\theta\|_\beta^{\beta c}} d\theta \\ &= m_1 + m_2. \end{aligned}$$

Obviously, as $|Z_i| \to +\infty$, m_1 has an exponential decreasing tail. Hence

$$\lim_{|Z_i|\to+\infty} \frac{m_1}{\pi(Z)} = 0.$$

By a change of variable $\theta = Z + y$, we have

$$m_2/\pi(Z) = \int\cdots\int_{\|Z+y\|_\beta>1} \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\|y\|^2} \frac{\|Z\|_\beta^{\beta c}}{\|Z+y\|_\beta^{\beta c}} dy.$$

To prove the theorem, it suffices to show the above expression converges to 1. In doing so, we shall apply the Dominated Convergence Theorem to show that we may pass the limit inside the above integral. After passing the limit, it is obvious that the integral becomes one.

The only argument left is to show that the Dominated Convergence Theoren can be applied. To do so, we seek an upper bound $F(y)$ for

$$\|Z\|_\beta^{\beta c}/\|Z+y\|_\beta^{\beta c} \text{ when } \|Z+y\|_\beta > 1.$$

Now for $\|Z+y\|_\beta > 1$,

$$\|Z\|_\beta^{\beta c} \le C_p(\|Z+y\|_\beta^{\beta c} + \|y\|_\beta^{\beta c}).$$

i.e.,

$$\frac{\|Z\|_\beta^{\beta c}}{\|Z+y\|_\beta^{\beta c}} \le C_p\left(1+\frac{\|y\|_\beta^{\beta c}}{\|Z+y\|_\beta^{\beta c}}\right) \le C_p(1+\|y\|_\beta^{\beta c}).$$

Hence if we take $C_p(1+\|y\|_\beta^{\beta c})$ as $F(y)$ then

$$\int\cdots\int_{\|Z+y\|_\beta>1} \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\|y\|^2} F(y)dy < +\infty.$$

Consequently, we may apply the Dominated Convergence Theorem, which completes the proof.

Proof of Theorem 1.5. Before relating to model (1.21), we shall work on the canonical form:

$$Z_i = \theta_i + \sigma\varepsilon_i, \ i = 1, 2, \ldots, d$$

where $\sigma > 0$, and ε_i are independently identically distribution standard normal random errors. Here $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_d)$ denotes the estimator in (1.10) with a defined in (1.22). For the rest of the paper C denotes a generic quantity independent of d and the unknown parameters. Hence the C's below are not necessarily identical. We shall first prove Lemma 1.2 below. Inequality (1.24) will be applied to the lower resolutions in the wavelet regression. The other two inequalities (1.25) and (1.26) are for higher resolutions.

Lemma 1.2 *For any $0 \le \beta < 2$, $0 < \delta < 1$, and some $C > 0$, independent of d and θ_i's, we have*

$$\sum_{i=1}^{d} E(\widehat{\theta}_i - \theta_i)^2 \le C\sigma^2 d(\ln d)^{(2-\beta)/2}, \tag{1.24}$$

and

$$E(\widehat{\theta}_i - \theta_i)^2 \le C(\theta_i^2 + \sigma^2 d^{\delta-1}(\ln d)^{-1/2}) \ \textit{if} \ \sum_1^d |\theta_i|^\beta \le \sigma^\beta \left(\frac{2-\beta}{2\beta}\right)^\beta \delta^2 m_\beta d. \tag{1.25}$$

Here and below, m_β denotes the expectation of $|\varepsilon_i|^\beta$, defined right above the statement of Theorem 1.5. Furthermore, for any $0 \le \beta < 1$, there exists $C > 0$ such that

$$E(\widehat{\theta}_i^A - \theta_i)^2 \le C\sigma^2 \ln d. \tag{1.26}$$

Proof: Proof: Without loss of generality, we will prove the theorem for the case $\sigma = 1$. By Stein's identity,

$$\begin{aligned} E & \quad (\widehat{\theta}_i - \theta_i)^2 \qquad (1.27) \\ &= E\left[1 + (Z_i - 2)I_i + \left(\frac{a^2|Z_i|^{2\beta-2}}{D} - 2a(\beta-1)\frac{|Z_i|^{\beta-2}}{D} + 2a\beta\frac{|Z_i|^{2\beta-2}}{D^2}\right)I_i^c\right] \end{aligned}$$

Here I_i denotes the indicator function $I(a|Z_i|^{\beta-2} > D)$ and $I_i^c = 1 - I_i$. Consequently

$$I_i = 1 \quad \text{if} \quad |Z_i|^{2-\beta} < a/D, \tag{1.28}$$

and

$$I_i^c = 1 \quad \text{if} \quad a|Z_i|^{\beta-2}/D \le 1. \tag{1.29}$$

¿From (1.27), and after some straightforward calculations,

$$\begin{aligned} E & \quad \sum_{i=1}^d (\widehat{\theta}_i - \theta_i)^2 \qquad (1.30) \\ &= d + E\left[\sum_{i=1}^d (|Z_i|^{2-\beta}|Z_i|^\beta - 2)I_i + \frac{a|Z_i|^{\beta-2}}{D}\left(\frac{a|Z_i|^\beta}{D} - 2(\beta-1) - 2\beta\frac{|Z_i|^\beta}{D}\right)I_i^c\right] \end{aligned}$$

Using this and the upper bounds in (1.28) and (1.29), we conclude that (1.30) is bounded above by

$$d + E\left[\sum_{i=1}^d \frac{a|Z_i|^\beta}{D} + \frac{a|Z_i|^\beta}{D} + 2\beta\frac{|Z_i|^\beta}{D}\right] + 2|\beta - 1|d \le C(\ln d)^{(2-\beta)/2}d,$$

completing the proof of (1.24).

To derive (1.25) for $1 < \beta < 2$, note that

$$E(1 + (Z_i^2 - 2)I_i) = \theta_i^2 + E(-Z_i^2 + 2)I_i^c.$$

This and (1.27) imply that

$$\begin{aligned} E(\widehat{\theta}_i - \theta_i)^2 &= \theta_i^2 + E\left\{\left[\left(\frac{a|Z_i|^{\beta-2}}{D}\right)^2 Z_i^2 - Z_i^2\right] I_i^c\right\} \\ &= +E\left\{\left[-2(\beta-1)\frac{a|Z_i|^{\beta-2}}{D} + 2\right] I_i^c\right\} - E\left[\left(2\beta a\frac{|Z_i|^{\beta-2}}{D}\ \frac{|Z_i|^{\beta}}{D}\right) I_i^c\right]. \end{aligned}$$

Using (1.30), one can establish that the last expression is bounded above by

$$\theta_i^2 + E[(-2(\beta-1)+2)I_i^c] + E2\beta\frac{|Z_i|^\beta}{D}I_i^c \le \theta_i^2 + E[(4+2\beta)I_i^c] \le \theta_i^2 + 8EI_i^c. \quad (1.31)$$

We shall show, under the condition in (1.25), that

$$EI_i^c \le C(|\theta_i|^2 + d^{\delta-1}(\log d)^{-1/2}). \quad (1.32)$$

This and (1.31) obviously establish (1.25). To prove (1.32), we shall consider two cases: (i) $0 \le \beta \le 1$ and (ii) $1 < \beta < 2$. For case (i), note that, for any $\delta > 0$, EI_i^c equals

$$\begin{aligned} P\left(a|Z_i|^{\beta-2} \le D\right) &= P\left(D \ge a|Z_i|^{\beta-2}, |Z_i| \le (2\ln d)^{1/2}/(1+\delta)\right) \\ &\quad + P\left(D \ge a|Z_i|^{\beta-2}, |Z_i| \ge (2\ln d)^{1/2}/(1+\delta)\right) \end{aligned}$$

Obviously, the last expression is bounded above by

$$P(D \ge (1+\delta)^{2-\beta} d m_\beta) + P(|Z_i| \ge (2\ln d)^{1/2}/(1+\delta)). \quad (1.33)$$

Now the second term is bounded above by

$$C(|\theta_i|^2 + (d^{1-\delta}\sqrt{\ln d})^{-1}) \quad (1.34)$$

by a result in Donoho and Johnstone (1994). To find an upper bound for the first term in (1.33), note that by a simple calculus

$$|Z_i|^\beta \le |\varepsilon_i|^\beta + |\theta_i|^\beta$$

due to $0 \le \beta \le 1$. Hence the first term of (1.33) is bounded above by

$$P\left(\sum_1^d |\varepsilon_i|^\beta \ge (1+\delta)^{2-\beta} dm_\beta - \sum |\theta_i|^\beta\right).$$

Replacing $\sum |\theta_i|^\beta$ by the assumed upper bound in (1.25), the last displayed expression is bounded above by

$$P\left(\sum_1^d |\varepsilon_i|^\beta \ge dm_\beta[(1+\delta)^{2-\beta} - (2-\beta)\delta^2]\right). \tag{1.35}$$

Using the inequality

$$(1+\delta)^{2-\beta} > 1 + (2-\beta)\delta,$$

one concludes that the quantity inside the bracket, is bounded below by

$$1 + (2-\beta)(\delta - \delta^2) > 1.$$

Hence the probability (1.35) decays exponentially fast. This and (1.34) then establish (1.32) for $0 \le \beta \le 1$.

To complete the proof for (1.25), all we need to do is to prove (1.32) for case (ii), $1 < \beta < 2$.

Similar to the argument for case (i), all we need to do is to show that the first term in (1.33) is bounded by (1.34). Now applying the triangle inequality

$$D^{1/\beta} \le \left(\sum |\varepsilon_i|^\beta\right)^{1/\beta} + \left(\sum |\theta_i|^\beta\right)^{1/\beta}$$

to the first term of (1.33) and using some straightforward algebraic manipulation, we obtain

$$\begin{aligned} & P\left(D \ge (1+\delta)^{2-\beta} dm_\beta\right) \\ \le\ & P\left(\sum_1^d |\varepsilon_i|^\beta \ge dm_\beta \left[\left\{(1+\delta)^{(2-\beta)/\beta} - \left(\frac{2-\beta}{2\beta}\right)\delta^{2/\beta}\right\}^\beta\right]\right). \end{aligned} \tag{1.36}$$

Note that

$$(1+\delta)^{(2-\beta)/\beta} \geq 1 + \frac{(2-\beta)\delta}{2\beta}$$

and consequently the quantity inside the bracket is bounded below by

$$\left[1 + \frac{2-\beta}{2\beta}(\delta - \delta^{2/\beta})\right]^{\beta} \geq 1 + (2-\beta)(\delta - \delta^{2/\beta})/2 > 1.$$

Now this shows that the probability on the right hand side decreases exponentially fast. Hence inequality (1.32) is established for case (ii) and the proof for (1.25) is now completed.

To prove (1.26) for $0 \leq \beta \leq 1$, we may rewrite (1.27) as

$$\begin{aligned} E\left(\widehat{\theta}_i - \theta_i\right)^2 = 1 \quad & + E(Z_i^2 - 2)I_i + E\left(|Z_i|^{2\beta-2}\left(\frac{a^2}{D^2} + \frac{2\beta a}{aD^2}\right) I_i^c\right) \\ & + 2(1-\beta)E\left[\frac{|Z_i|^{\beta-2}a}{D} I_i^c\right] \end{aligned} \tag{1.37}$$

The inequality (1.26), sharper than (1.24), can be possibly established due to the critical assumption $\beta \leq 1$, which implies that

$$|Z_i|^{2\beta-2} < \left(\frac{a}{D}\right)^{-(2-2\beta)/(2-\beta)} \quad \text{if} \quad I_i^c = 1. \tag{1.38}$$

Note that the last term in (1.37) is obviously bounded above by $2(1-\beta)$. Furthermore, replace $|Z_i|^{2\beta-2}$ in the third term on the right hand side of (1.37) by the upper bound in (1.38) and replace Z_i^2 in the second term by the upper bound below

$$|Z_i|^2 < (a/D)^{2/(2-\beta)} \text{ when } I_i = 1,$$

which follows easily for (1.28). We then obtain an upper bound for (1.37)

$$\begin{aligned} & 1 + E\left(a/D\right)^{2/(2-\beta)} + E\left[(a/D)^{(2\beta-2)/(2-\beta)}\left(\frac{a^2}{D^2} + 2\frac{\beta a}{D^2}\right) I_i^c\right] + 2(1-\beta) \\ \leq \quad & (3-2\beta) + CE(a/D)^{2/(2-\beta)}. \end{aligned}$$

Here, in the last inequality, $2\beta a/D^2$ was replaced by $2\beta a^2/D^2$. To establish (1.26), obviously the only thing left to do is

$$E(a/D)^{2/(2-\beta)} \leq C \ln(d). \tag{1.39}$$

This inequality can be established if we can show that

$$E(d/D)^{2/(2-\beta)} \leq C \tag{1.40}$$

since the definition of a and a simple calculation show that

$$a^{2/(2-\beta)} = Ca^{2/(2-\beta)} \ln(d).$$

To prove (1.40), we apply Anderson's theorem (Anderson 1955) which implies that $|Z_i|$ is stochastically larger than $|\varepsilon_i|$. Hence

$$E(d/D)^{2/(2-\beta)} \leq E\left[d/\left(\sum |\varepsilon_i|^\beta\right)\right]^{2/(2-\beta)},$$

which is bounded by $A + B$. Here

$$A = E\left[d/\left(\sum |\varepsilon_i|^\beta\right)\right]^{2/(2-\beta)} I\left(\sum_1^d |\varepsilon_i|^\beta \leq dm_\beta/2\right)$$

and

$$B = E\left[d/\left(\sum |\varepsilon_i|^\beta\right)\right]^{2/(2-\beta)} I\left(\sum_1^d |\varepsilon_i|^\beta > dm_\beta/2\right)$$

and as before $I(\cdot)$ denotes the indicator function.

Now B is obviously bounded above by

$$(2/m_\beta)^{2/(2-\beta)} < C.$$

Also by Cauchy–Schwartz inequality

$$A^2 \leq E\left[d/\left(\sum |\varepsilon_i|^\beta\right)\right]^{4/(2-\beta)} P\left(\sum_1^d |\varepsilon_i|^\beta \leq dm_\beta/2\right) < C.$$

Here the last inequality holds since the probabiity decays exponentially fast. This completes the proof for (1.40) and consequently for (1.26).

Now we apply Lemma 1.2 to the wavelet regression. Equivalently we shall consider the model

$$Z_{jk} = \theta_{jk} + \varepsilon_{jk}/\sqrt{n}, \ k = 1, \ldots, 2^j, \tag{1.41}$$

where θ_{jk}'s are wavelet coefficients of function f, and ε_{jk}'s are i.i.d. standard normal random variables. For the details of reasoning supporting the above statement, see, for example, Section 9.2 of Cai (1999), following the ideas of Donoho and Johnstone (1997 and 1998). Also assume that θ's live in the Besov space $B^{\alpha}_{p,q}(M)$ with smoothness α and shape parameters p and q, i.e.,

$$\sum_j 2^{jq(\alpha+1/2-1/p)} \left(\sum_k |\theta_{jk}|^p \right)^{q/p} \leq M^q \tag{1.42}$$

for some positive constants α, p, q and M. The estimator $\widehat{\theta}$ below for model (1.41) refers to (1.20) with a defined in (1.22) and $\sigma^2 = 1/n$. For such a $\widehat{\theta}$, the total risk can be decomposed into the sum of the following three quantities:

$$R_1 = \sum_{j<j_0} \sum_k E(\widehat{\theta}_{jk} - \theta_{jk})^2,$$

$$R_2 = \sum_{J>j\geq j_0} \sum_k E(\widehat{\theta}_{jk} - \theta_{jk})^2$$

and

$$R_3 = \sum_{j\geq J} \sum_k E(\widehat{\theta}_{jk} - \theta_{jk})^2$$

where $j_0 = [\log_2(C_\delta n^{1/(2\alpha+1)})]$, and C_δ is a positive constant to be specified later. Applying (1.24) to R_1, which corresponds to the risk of low resolutions, we establish some simple calculation

$$R_1 \leq C(\ln n)^{(2-\beta)/2} n^{-2\alpha/(2\alpha+1)}. \tag{1.43}$$

For $j \geq j_0$, (1.42) implies

$$\sum_k |\theta_{jk}|^p \leq M^p 2^{-jp(\alpha+1/2-1/p)} = M^p 2^j 2^{-jp(\alpha+1/2)}. \qquad (1.44)$$

Furthermore, for $p \geq \beta$

$$2^{-jp(\alpha+1/2)} \leq 2^{-j\beta(\alpha+1/2)} \leq 2^{-j_0\beta(\alpha+1/2)} = (C_\delta)^{-\beta(\alpha+1/2)}\sigma^\beta.$$

Choose $C_\delta > 0$ such that

$$M^p / C_\delta^{(1/2+\alpha)\beta} = \left(\frac{2-\beta}{2\beta}\right)^\beta \left(\frac{1}{2\alpha+1}\right)^2 m_\beta.$$

This then implies that

$$\begin{aligned}\sum_k |\theta_{jk}|^p &\leq \frac{M^p}{C_\delta^{(1/2+\alpha)\beta}} 2^j \sigma^\beta \\ &\leq \left(\frac{2-\beta}{2\beta}\right)^\beta \left(\frac{1}{2\alpha+1}\right) m_\beta 2^j \sigma^\beta,\end{aligned}$$

satisfying the condition in (1.25) for $d = 2^j$ and $\delta = (2\alpha+1)^{-1}$.

Now for $p \geq 2$ we give an upper bound for the total risk.

¿From (1.25), we obtain

$$R_2 + R_3 \leq C \sum_{j \geq j_0} \sum_k \theta_{jk}^2 + o(n^{-2\alpha/(2\alpha+1)})$$

and from Hölder inequality the first term is bounded above by

$$\sum_{j \geq j_0} 2^{j(1-2/p)} \left(\sum_k |\theta_{jk}|^p\right)^{2/p}.$$

Then inequality (1.44) gives

$$\begin{aligned}R_2 + R_3 &\leq C \sum_{j \geq j_0} 2^{j(1-2/p)} 2^{-j2(\alpha+1/2-1/p)} + o(n^{-2\alpha/(2\alpha+1)}) \\ &\leq C \sum_{j \geq j_0} 2^{-j2\alpha} + o\left(n^{-2\alpha/(2\alpha+1)}\right) \\ &\leq Cn^{-2\alpha/(2\alpha+1)}.\end{aligned}$$

This and (1.43) imply (1.23) for $0 \le \beta \le 2$ and $p \ge 2$.

Note that for $\beta = 2$, the proof can be found in Donoho and Johnstone (1995). For $\beta \ne 2$, our proof is very different and much more involved.

To complete the proof of the theorem, we now focus on the case $0 \le \beta \le 1$, and $2 > p \ge \max\{1/\alpha, \beta\}$ and establish (1.23). We similarly decompose the risk of $\widehat{\theta}$ as the sum of R_1, R_2 and R_3. Note that the bound for R_1 in (1.43) is still valid. Inequalities (1.25) and (1.26) imply

$$R_2 \le \sum_{J \ge j \ge j_0} \sum_k \theta_{jk}^2 \wedge \frac{\log n}{n} + o\left(\frac{1}{n^{1-\delta}}\right)$$

for some constants $C > 0$. Furthermore, the following inequality

$$\sum x_i \wedge A \le A^{1-t} \sum x_i^t,\ x_i \ge 0,\ A > 0,\ 1 \ge t > 0$$

implies

$$\sum_{J \ge j \ge j_0} \sum_k \theta_{jk}^2 \wedge \frac{\log n}{n} \le \left(\frac{\log n}{n}\right)^{1-p/2} \sum_{J > j \ge j_0} \sum_k |\theta_{jk}|^p.$$

Some simple calculations, using (1.44), establish

$$\begin{aligned} R_2 &\le C\left(\frac{\log n}{n}\right)^{1-p/2} \sum_{J > j \ge j_0} 2^{-jp(\alpha+1/2-1/p)} + o(n^{-2\alpha/(2\alpha+1)}) \\ &\le C(\log n)^{1-p/2} n^{-2\alpha/(2\alpha+1)} \end{aligned} \tag{1.45}$$

¿From Hölder inequality, it can be seen that R_3 is bounded above by

$$\sum_{j \ge j_0} \left(\sum_k |\theta_{jk}|^p\right)^{2/p}.$$

Similar to (1.45), we obtain the upper bound of R_3,

$$R_3 \le C \sum_{j \ge J} 2^{-j2(\alpha+1/2-1/p)} = o(n^{-2\alpha/(2\alpha+1)}),$$

where J is taken to be $\log_2 n$. Thus for $0 \le \beta \le 1$ nd $2 \ge p \ge \max\{1/\alpha, \beta\}$, we have

$$\sup_{f \in B_{p,q}^\alpha} E\|\widehat{\theta} - \theta\|^2 \le C(\log n)^{1-\beta/2} n^{-2\alpha/(2\alpha+1)}.$$

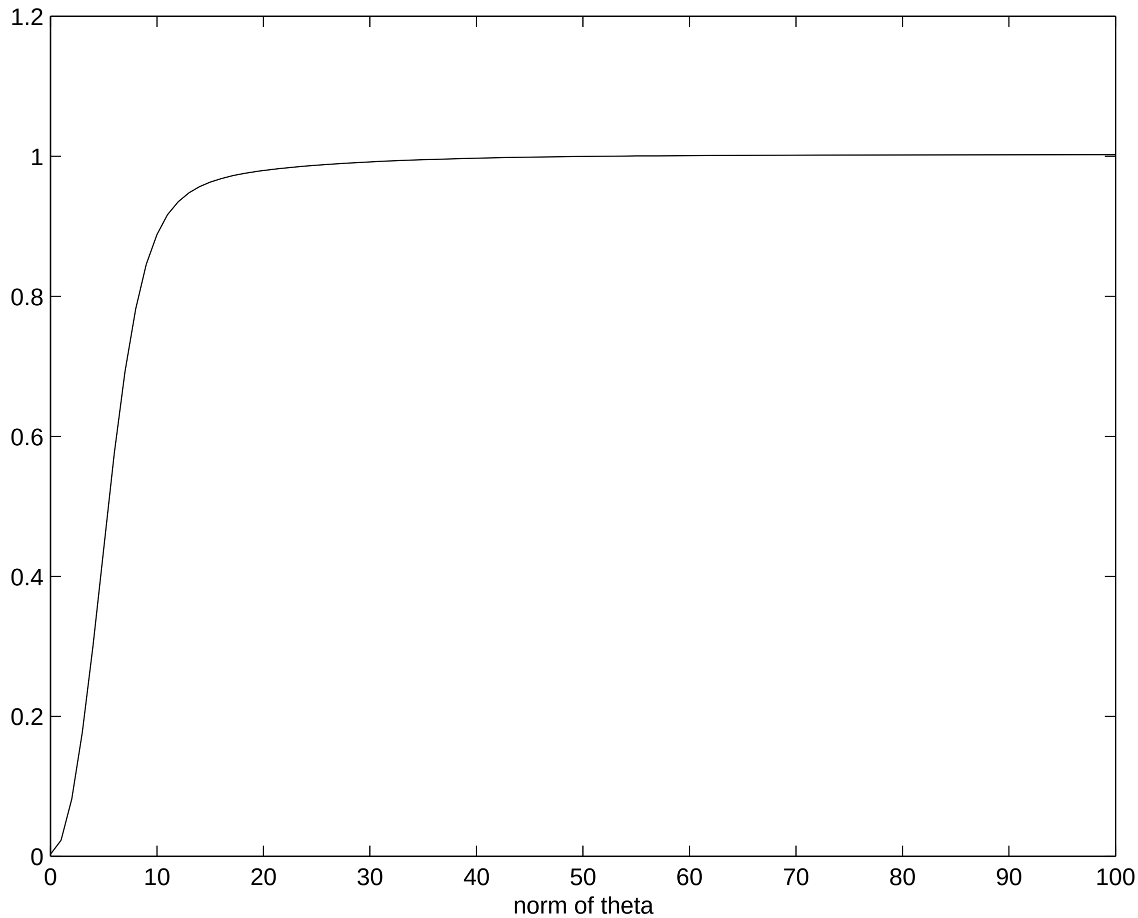

Figure 1.1: Relative Frequentist risk of the proposed estimator (1.12) to p=50.

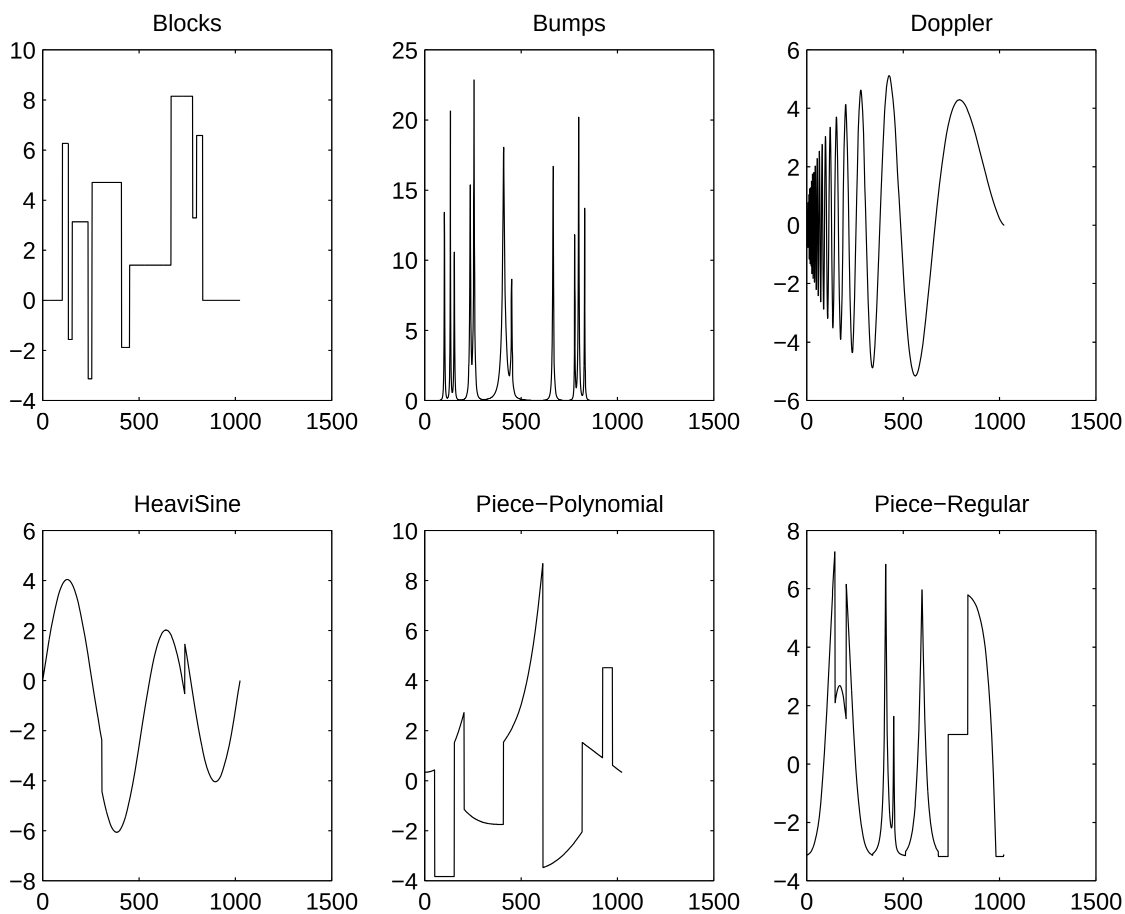

Figure 1.2: The curves represent the true curves f(t)in (1.18).

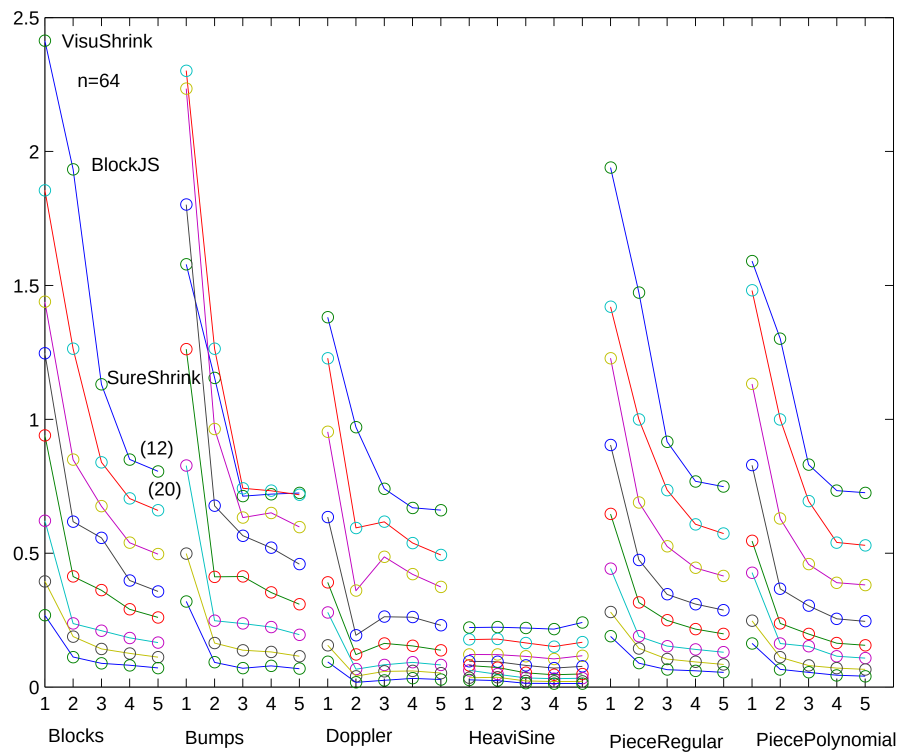

Figure 1.3: In each of the six casescorresponding to Blocks, Bumps, etc., the eight curves plot the risk function, from top to the bottom, when $n = 64, 128, \ldots, 8192$. For each curve (see for example, the top curve on the left), the circles "o" from left to the right give, with respect to n, the relative risks of VisuShrink, Block James–Stein, SureShrink, and the proposed methods (1.12) and (1.20).

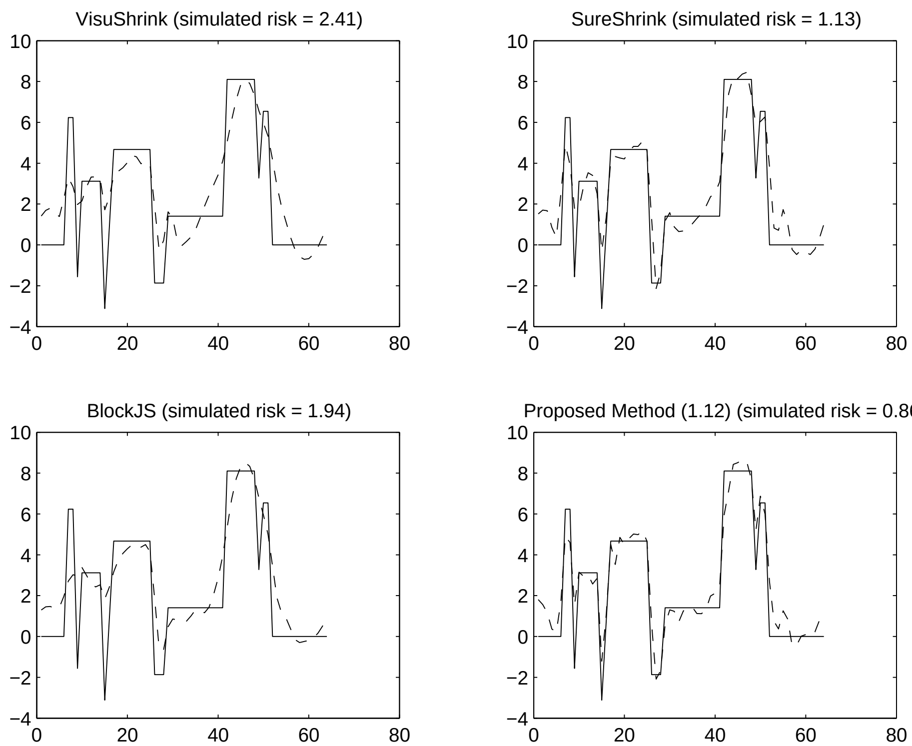

Figure 1.4: Solid lines represent the true curves, wheredotted lines represent the curves corresponding to various estimators. The simulated risk is based on 500 simulations.

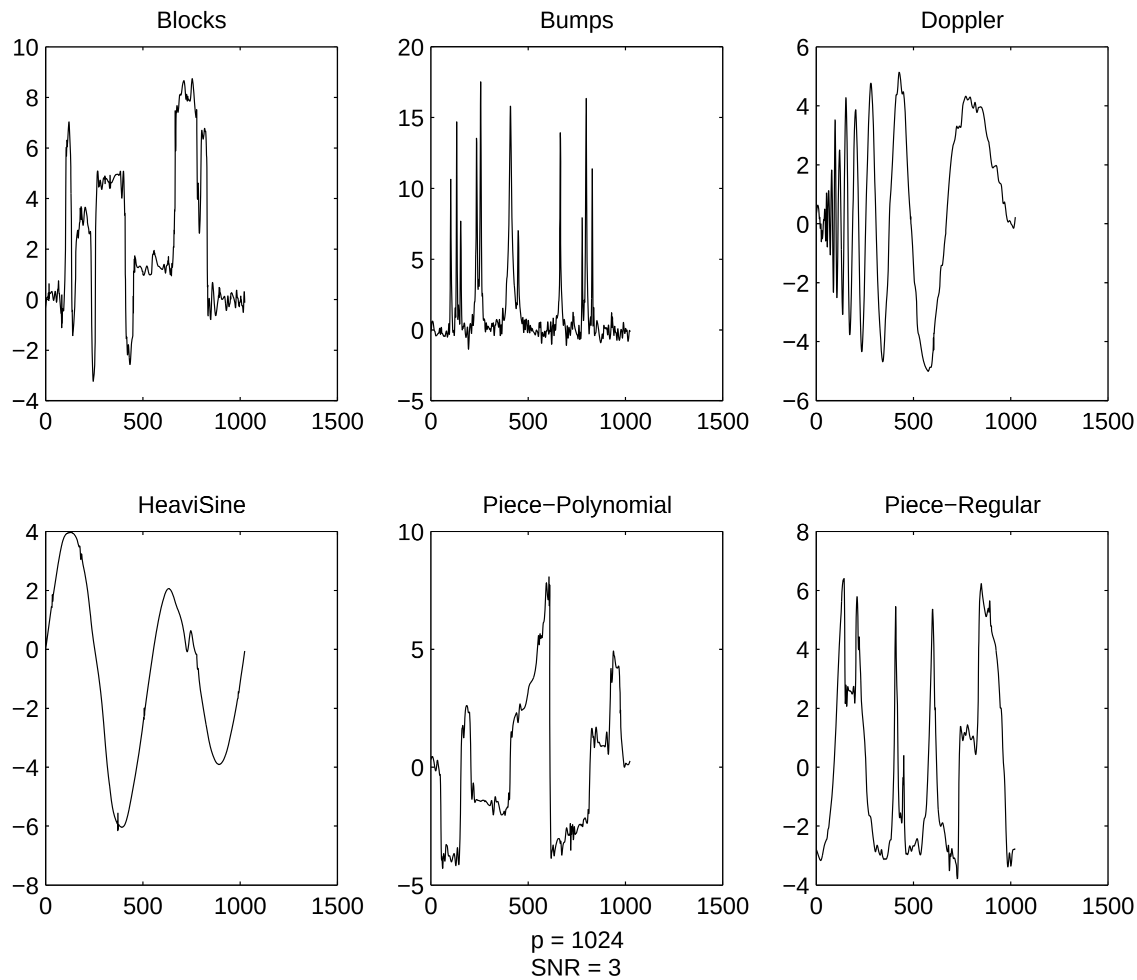

Figure 1.5: Proposed Estimator (1.12) Applied to Reconstruct Figure 1.2.

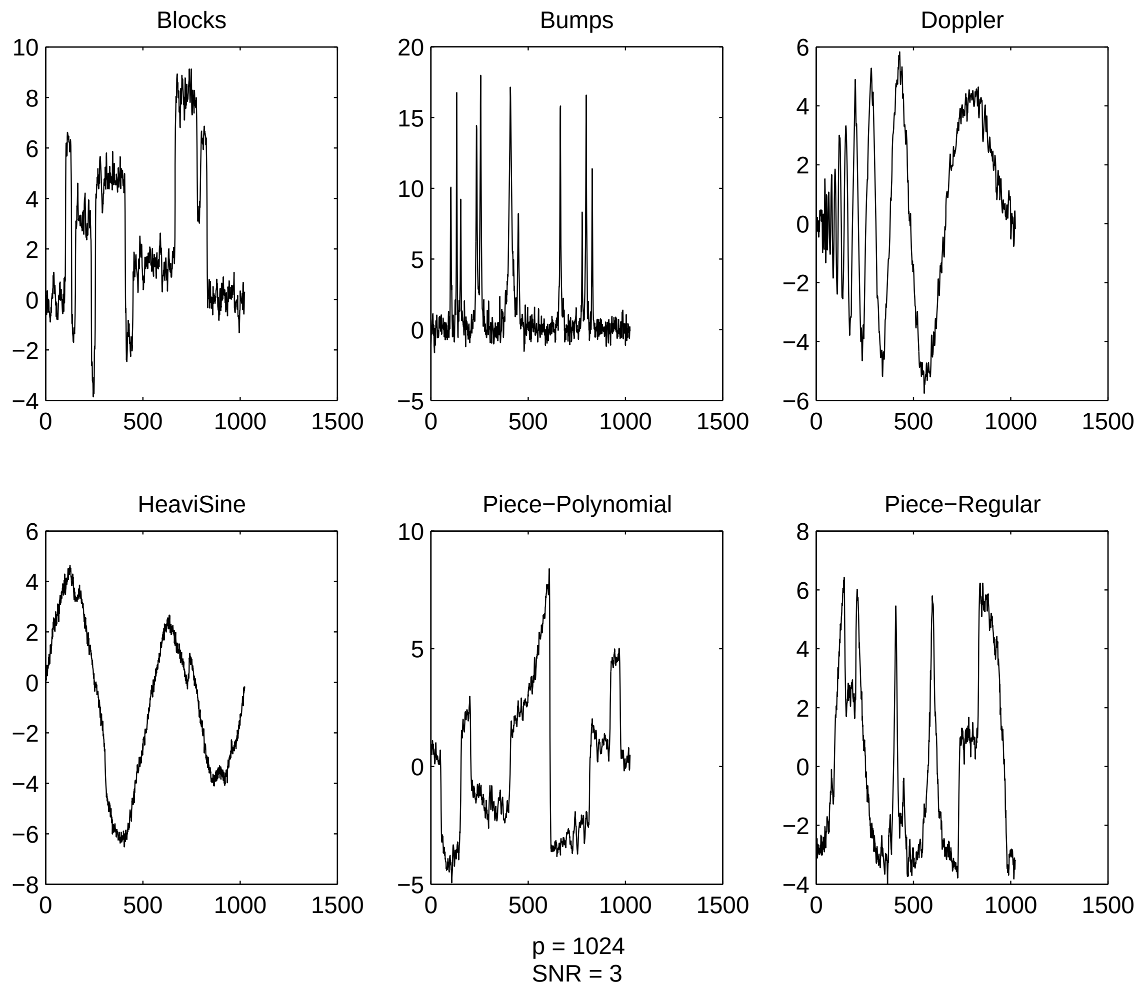

Figure 1.6: James-Stein Positive Part Applied to Reconstruct Figure 1.2.

BIBLIOGRAPHY

[1] Antoniadis and Fan (2001), Regularized wavelet approximations (with discussion), *J. Am. Statist. Ass.* **96**, 939–967.

[2] Antoniadis, Leporini and Desquet (2002), *Statistica Neerlandica* (to appear).

[3] Beran, R. and Dümbgen, L. (1998), Modulation of estimators and confidence set, *Ann. Statist.* **26**, 1826–1856.

[4] Berger, J. (1976), Tail minimaxity in location vector problems and its applications, *Ann. Statist.* **4**, 33–50.

[5] Berger, J. (1980), Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters, *Ann. Statist.* **8**, 545–571.

[6] Brown, L. D. (1971), Admissible estimators, recurrent diffusions, and insoluble boundary value problems, *Ann. Math. Statist.* **42**, 855–903.

[7] Cai, T. (1999), Adaptive wavelet estimation: A block thresholding and oracle inequality approach, *Ann. Statist.* **27**, 898–924.

[8] Donoho, D. L. and Johnstone, I. (1994), Ideal spatial adaption via wavelet shrinkage, *Biometrika* **81**, 425–455.

[9] Donoho, D. L. and Johnstone, I. (1995), Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Stat. Assoc.* **90**, 1200–1224.

[10] Gao, H. Y. (1998), Wavelet shrinkage denoising using non–negative garrote, *J. Comput. Graph. Statist.* **7**, 469–488.

[11] Gauch, H. (1993), Prediction, parsimony and noise, *American Scientist* **81**, 468–478.

[12] George, E. I. (1986a), Minimax multiple shrinkage estimation, *Ann. Statist.* **14**, 188–205.

[13] George, E. I. (1986b), Combining minimax shrinkage estimation, *J. Am. Statist. Ass.* **81**, 437–445.

[14] James, W. and Stein, C. (1961), Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 311–319.

[15] Lehmann, E. L. (1983), *Theory of Point Estimation,* Wiley, New York.

[16] Lehmann, E. L. and Casella, G. C. (1998), *Theory of Point Estimation,* Second edition, Springer-Verlag, New York.

[17] Mallat, S. G. (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. on Patt. Anal. Mach. Intell.* **11**(7), 674–693.

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

[19] Vidakovic, B. (1999), Statistical Modeling by Wavelets, John Wiley & Sons, Inc., New York.

[20] Zhou, H. H. and Hwang, J. T. G. (2003), Minimax estimation with thresholding, *Cornell Statistical Center Technical Report.*

Chapter 2

Asymptotic Equivalence Theory of the Gaussian Variance Regression Experiment

2.1 Introduction

One of the most important statistical contributions of Lucien Le Cam is the asymptotic equivalence theory. A basic principle of asymptotic equivalence theory is to approximate general statistical models (also called experiments) by simple ones. This idea can be traced back to the paper of Wald (1943). Wald wanted to reduce the general problem to a simpler multivariate normal case. But Wald's demands on a set-to-set relation in the space of n observations to one in the space of maximum likelihood estimates are too strict and can be satisfied only under special circumstance. However the idea to approximate some experiments by simpler ones is illuminating. In 1950s, David Blackwell and Stein proved a result which gives a necessary and sufficient condition such that one statistical experiment is "better" or "more informative" than the other experiment, followed by an unpublished report in Rand Memorandum by Bohnenblust, Shapley and Sherman (1949), which is now called Blackwell-Sherman-Stein theorem. Le Cam (1964) introduced his deficiency distance between two models motivated by the Blackwell-Sherman-Stein theorem. The deficiency distance can be seen as an attempt to make Blackwell-Sherman-Stein theorem quantitative, i.e., to quantify the "deficiency" between two experiments. This extension enables him to address Wald's problem of reduction of the general problem to a simpler one.

A great success of Le Cam's deficiency distance has been achieved in the classical parametric setting where asymptotic questions very often can be efficiently reduced to appropriate questions involving only normal distributions. Le Cam (1972) consider *weak convergence to Gaussian shift experiments* defined by his deficiency distance under the classical local asymptotic normality conditions (LAN). Two of the most valuable results of that theory are Hájek-Le Cam asymptotic minimax and convolution theorems (see Le Cam (1972)). The Hájek-Le Cam asymptotic minimax and convolution theorems are widely used in asymptotic statistics. In particular, it is well known that they saved the Cramér-Rao bound from the superefficiency by studying a minimax risk. But the story they can tell is far from complete. One limitation of Hájek-Le Cam's results is that they often require $n^{-1/2}-$ consistency, which is impossible in most of infinite-dimensional situations. Of course Le Cam's results covers certain infinite-dimensional situations, but his main focus appears to be just the finite-dimensional models. Another important limitation is that their results can only yields lower bounds for the risk of estimates. They can not yield upper bound. In most of literature the efficiency of procedures is more or less shown on an ad hoc basis. However convergence in the strong sense of the Le Cam distance Δ can yield both lower and upper bounds. ‘ A breakthrough, away from these two limitations in the last paragraph, is a result by Nussbaum in 1996 following the work of Brown and Low (1996). Let Δ be Le Cam's deficiency pseudodistance between experiments having the same parameter space. For two sequences of experiments E_n and F_n we shall say that they are *asymptotically equivalent,* if $\Delta(E_n, F_n) \to 0$ as $n \to \infty$. Nussbaum established the global asymptotic equivalence of the white-noise problem to the nonparametric density problem under a Lipschitz smoothness condition. The significance of asymptotic equivalence is that all asymptotically optimal statistical procedures can be carried over from one experiment to the other for bounded loss functions (or for unbounded ones, by truncation). The Gaussian case is well studied. A lot of results are known in minimax estimation for the Gaussian case such as optimal convergence rates and optimal constants (see Donoho, Johnstone, Kerkyacharian, and Picard (1995), Tsybakov(1997), and references therein). Using Nussbaum's theorem, one can readily transfer all these Gaussian results to the case of density estimation.

There have been several developments in the past years. Grama and Nussbaum (1998, 2003a) generalize the global asymptotic result for density estimation in Nussbaum (1996) to a general nonparametric regression under Lipschitz smoothness conditions, which includes generalized linear models and location type regression. Milstein and Nussbaum (1998) showed that some diffusion problems can be approximated by their discrete versions, which can be seen as an extension of Brown and Low (1996). In Brown, Cai, Low and Zhang (2002) the asymptotic equivalence for nonparametric regression between random design and the fixed design was shown.

Recently Brown, Carter, Low and Zhang (2004) made an important progress in this direction. they extended the result of Nussbaum (1996) to the Besov smoothness conditions which includes the Lipschitz classes with any smoothness $\alpha > 1/2$. In this paper, we establish the global asymptotic equivalence between the Gaussian variance regression experiment and the Gaussian white noise experiment under the Besov smoothness constraints including some discontinuous functions. A closely related work is Grama and Nussbaum (1998) in which they proved the result under the Lipschitz smoothness assumption on unknown function. In Brown, Carter, Low and Zhang (2004) their proof is constructive, i.e., the asymptotic equivalence is established by constructing explicit mappings between these two experiments. But the approach of Brown, Carter, Low and Zhang (2004) can not be applied directly in the Gaussian variance regression problem. Our proof here is nonconstructive; however, the exploring of asymptotic equivalence under Besov smoothness conditions is appropriate to transfer recent results in the Gaussian case (see Donoho and Johnstone (1995)) to its equivalent models. One could envision the results in Grama and Nussbaum (1998, 2003a) can be proved to stand under Besov smoothness conditions by similar techniques proposed in this paper.

The approach of Brown, Carter, Low and Zhang (2004) can not be applied directly here, because they implicitly used the fact that the mean of all X_j with $X_i \overset{i.i.d.}{\sim} Bin(1,p)$ is an efficient estimator of p. A parametric model related to the experiment $E_{1,n}$ is the location problem

$$z_j = \theta + \log \xi_j^2,\ \xi_j \text{ i.i.d. } N\left(0,1\right), \theta \in \mathbb{R} \text{ unkown},\ j = 1, \ldots, n.$$

In this model the mean of all z_j is not an efficient estimator of θ, which motivates us to use the likelihood ratio to establish the Gaussian approximation in a parameter-local framework, similar to Nussbaum (1996) and Grama and Nussbaum (1998, 2003a). In section 2.2, local expansions of the likelihood ratios for both experiments are given. Then we expand the approximation terms in the log-likelihood ratios by Haar wavelets. This scheme is different from that of Nussbaum (1996) or Grama and Nussbaum (1998, 2003a), where they applied functional Komlós-Major-Tusnády inequality in Grama and Nussbaum (2003b), so that they could only establish the asymptotic equivalence under Lipschitz smoothness conditions. Thus we could apply a more delicate coupling method with the help of Haar expansions. A multiresolution coupling methodology (similar to the Hungarian construction) is then used to establish asymptotic equivalence (see also Carter (2002), Brown, Carter, Low and Zhang (2004)) in the section 2.3. For each resolution, our coupling approach is more elegant than the traditional quantile coupling methods; essentially we just use quantile couplings between independent beta's and independent normals. The quantile coupling method was introduced in Komlós, Major, and Tusnády(1975), which is considered one of the most important probability papers of the last forty years, and now often referred to as the KMT paper. Their coupling greatly simplifies the derivation of many classical results (see Shorack & Wellner (1986)), and is a powerful way to obtain strong approximation results. For the quantile coupling between a B_n distributed Beta $(n/2, n/2)$ and a normal random variable, in section 2.6 we establish

$$\left|nB_n - n/2 - \left(n^{1/2}/2\right) Z\right| \leq \frac{C}{\sqrt{n}} \left(1 + |Z|^3\right), when \ |Z| \leq \varepsilon\sqrt{n},$$

where Z is a standard normal, and $C, \varepsilon > 0$ do not depend on n. The right hand side of the equation above improves the classical bound $C\left(1 + |Z|^2\right)$ with a rate $1/\sqrt{n}$, which is of independent interest. In particular this sharp bound is helpful for us to establish the asymptotic equivalence under the Besov smoothness conditions. The main proof of local asymptotic equivalence is given in section 2.4. In section 2.5 we construct the preliminary estimators with appropriate convergence rates for both experiments such that the local approximations can be glued together in a global approximation. Some technical lemmas are given in the appendix, section 2.7.

To formulate our result, define a basic parameter space Σ as follows. For a given $\epsilon > 0$, we define a set $\mathcal{F}_\epsilon$ as the set of densities on $[0, 1]$ bounded below by ϵ

and above by ϵ^{-1}:

$$\mathcal{F}_{\epsilon}=\left\{f:\int_0^1 f=1,\epsilon\leq f(x)\leq\epsilon^{-1},\ x\in[0,1]\right\}. \tag{2.1}$$

The parameter space Σ is a subset of $\mathcal{F}_{\epsilon}$, and contained in Besov balls $B^{\alpha}_{2,2}(M)$ and $B^{\alpha}_{6,6}(M)$ with $\alpha>1/2$ and $M>0$, i.e.,

$$\Sigma\subset\mathcal{F}_{\epsilon}\cap B^{\alpha}_{2,2}(M)\cap B^{\alpha}_{6,6}(M)\text{, for }\epsilon>0,\ \alpha>1/2\text{ and }M>0. \tag{2.2}$$

Now we review the Haar basis to give a definition of the Besov balls (see also Brown,Carter, Low and Zhang (2004)). Let I_{kl} be the indicator functions of intervals $2^{-k}(l-1,l]$ and define

$$\begin{aligned}\phi_{kl}(x)&=2^{k/2}I_{kl},\\ \psi_{kl}(x)&=2^{k/2}(I_{k+1,2l-1}-I_{k+1,2l}),\end{aligned}$$

where $k\geq 0$ and $1\leq l\leq 2^k$, i.e., ψ_{kl} are the orthonormal Haar wavelets. Let

$$f_{kl}=\langle f,\phi_{kl}\rangle,\ \widetilde{f}_{kl}=\langle f,\psi_{kl}\rangle, \tag{2.3}$$

be the Haar coefficients of $f\in L^2([0,1])$. For $p\geq 1$, $q\geq 1$ and $1>\alpha\geq 1/p-1/2$, the Besov sequence norms $\|f\|^{\alpha}_{p,q}$ for the Haar coefficients are given by

$$\|f\|^{\alpha}_{p,q}=\left\{|f_{01}|^q+\sum_{k=0}^{\infty}\left[2^{k(\alpha+1/2-1/p)}\left(\sum_{l=1}^{2^k}\left|\widetilde{f}_{kl}\right|^p\right)^{1/p}\right]^q\right\}^{1/q}, \tag{2.4}$$

and for $M>0$ define Besov balls

$$B^{\alpha}_{p,q}(M)=\left\{f:\|f\|^{\alpha}_{p,q}\leq M\right\}. \tag{2.5}$$

Remark 2.1 *From the definition of the Besov norm in (2.4) and Besov balls above, the condition $\alpha>1/p$ doesn't necessarily implies all functions in $B^{\alpha}_{p,q}(M)$*

are in Hölder space $C^{\alpha-1/p}$. *For instance, the indicator function* $2I_{(0,1/2]}$ *is in* $B^{\alpha}_{p,q}(M)$ *for any* α, p *and* q, *but it is not continuous on* $[0,1]$, *then not contained in* $C^{\alpha-1/p}$. *The definition of Besov balls via sequence spaces is more general than that via modulus.*

Remark 2.2 *Lemma 2.8 in the appendix shows that*

$$\{\log f : f \in \Sigma\} \subset B^{\alpha}_{2,2}(M') \cap B^{\alpha}_{6,6}(M'), \text{ for some } M' > 0.$$

This property will be used very often.

The following theorem is our main result.

Theorem 2.1 *Let the parameter space* Σ *be defined above. The experiments given by observations*

$$E_{1,n} \;:\; z(j) = J_{j,n}(f)\,\xi_j^2, \xi_j \text{ i.i.d. } N(0,1), \; j = 1, \ldots, n \tag{2.6}$$

$$E_{2,n} \;:\; dy(t) = \frac{1}{\sqrt{2}} \log f(t)\, dt + n^{-1/2} dW(t), \tag{2.7}$$

where $J_{j,n}(f) = n \int_{(j-1)/n}^{j/n} f$ *and unknown* $f \in \Sigma$, *are asymptotically equivalent.*

2.2 The Structure of Likelihood

In this section we give the local stochastic expansions of the likelihood ratio for both experiments, when all functions f are centered around a fixed function f_0. Then we expand the approximation terms in the log-likelihood ratios by Haar wavelets. In the next two sections we will prove the local asymptotic equivalence of $E_{1,n}$ and $E_{2,n}$ by comparing the Haar expansions of the approximation terms in their log-likelihood ratios.

Recall that the experiment $E_{1,n}$ defined in (2.6) is given by observations

$$z(j) = J_{j,n}(f)\,\xi_j^2,$$

or equivalently,

$$z(j) = \log J_{j,n}(f) + \log \xi_j^2,$$

where ξ_j are i.i.d. $N(0,1)$, and $j = 1, \ldots, n$. Since

$$\log J_{j,n}(f) \approx J_{j,n}(\log f)\,,$$

lemma 2.10 in the appendix shows that the experiment $E_{1,n}$ is asymptotically equivalent to $E_{1,n}^1$ given by observations

$$z(j) = J_{j,n}(\log f) + \log \xi_j^2,\; j = 1, \ldots, n.$$

And lemma 2.12 in the appendix shows that the experiment $E_{2,n}$ is asymptotically equivalent to $E_{2,n}^1$ given by observations

$$y_j = \frac{1}{\sqrt{2}} J_{j,n}(\log f) + \eta_j, \eta_j \text{ i.i.d.} N(0,1)\,, j = 1, \ldots, n.$$

See also Brown and Low (1996) for the detail. We will show $\Delta\left(E_{1,n}^1, E_{2,n}^1\right) \to 0$, which gives $\Delta\left(E_{1,n}, E_{2,n}\right) \to 0$.

The asymptotic equivalence between $E_{1,n}^1$ and $E_{2,n}^1$ will be established in a parameter-local framework first. Let γ_n be the sequence

$$\gamma_n = (\log n)^{-2}\, n^{-1/4}$$

and for any $f_0 \in \Sigma$ define a class $\Sigma_n(f_0)$ by

$$\Sigma_n(f_0) = \left\{ f \in \Sigma; \|f - f_0\|_p \leq \gamma_n \right\}. \tag{2.8}$$

So we have large " nonparametric " neighborhoods which are attainable for estimators if Σ is in $B_{p,p}^{\alpha}$ with $\alpha > 1/2$ and $p \geq 1$ for both experiments $E_{1,n}$ and $E_{2,n}$

(see the section "The preliminary estimators" for the detail). In our setting the number p is 6. Since functions are supported on $[0,1]$ with Lebesgue measure 1, Jensen's inequality implies

$$\|f-f_0\|_p^p \le \gamma_n^p \text{ , for } 0<p\le 6 \text{ anf } f\in\Sigma_n(f_0)\,, \tag{2.9}$$

which will be used later.

In the local framework $f\in\Sigma_n(f_0)$, the experiment $E_{1,n}^1$ is equivalent to

$$z(j)=J_{j,n}\left(\log\left(f/f_0\right)\right)+\log\xi_j^2,\; j=1,\ldots,n.$$

Since $J_{j,n}\left(\log f/f_0\right)$ is small, we have

$$J_{j,n}\left(\log f/f_0\right)\approx-\log(1+J_{j,n}\left(\log\left(f_0/f\right)\right)).$$

Lemma 2.11 in the appendix shows that the experiment $E_{1,n}^1$ is locally asymptotically equivalent to $E_{1,n}^2$ given by observations

$$z(j)=-\log(1+J_{j,n}\left(\log\left(f_0/f\right)\right))+\log\xi_j^2,\; j=1,\ldots,n,$$

i.e.,

$$\Delta\left(E_{1,n}^1,E_{2,n}^1\right)\to 0\Longleftrightarrow\Delta\left(E_{1,n}^2,E_{2,n}^1\right)\to 0,\text{ for } f\in\Sigma_n\left(f_0\right).$$

The density of the random variable $\log\xi_i^2$ is

$$p\left(t\right)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\exp\left(t\right)}{2}\right)\exp\left(\frac{t}{2}\right),$$

then the likelihood process $\Lambda_{1,n}$ for the experiment $E_{1,n}^2$ is

$$\begin{aligned}\Lambda_{1,n} &= \prod_{j=1}^{n}\frac{p\left(\log\xi_i^2+\log(1+J_{j,n}\left(\log\left(f_0/f\right)\right))\right)}{p\left(\log\xi_i^2\right)} \\ &= \exp\begin{pmatrix}\sum_{j=1}^{n}\left(-\frac{1}{2}\left(\xi_j^2-1\right)\right)\left(J_{j,n}\left(g\right)\right)\\ -\frac{1}{2}\sum_{j=1}^{n}\left(J_{j,n}\left(g\right)-\log\left(1+J_{j,n}\left(g\right)\right)\right)\end{pmatrix},\end{aligned} \tag{2.10}$$

where $g = \log(f_0/f)$. Let $\zeta_j = -\left(\xi_j^2 - 1\right)/\sqrt{2}$ such that

$$E\zeta_j = 0,\ Var(\zeta_j) = 1.$$

Use an approximation

$$\left(J_{j,n}\left(g\right) - \log\left(1 + J_{j,n}\left(g\right)\right)\right) \approx \frac{1}{2} J_{j,n}^2\left(g\right),$$

then

$$\Lambda_{1,n} \approx \left(\sum_{j=1}^{n} \frac{1}{\sqrt{2}} \zeta_j J_{j,n}(g) - \frac{1}{4} \sum_{j=1}^{n} J_{j,n}^2\left(g\right)\right).$$

On the heuristic level, we would like to replace ζ_i by standard normal η_i to obtain a Gaussian likelihood ratio

$$\Lambda_{2,n} = \exp\left(\sum_{j=1}^{n} \frac{1}{\sqrt{2}} \eta_j J_{j,n}(g) - \frac{1}{4} \sum_{j=1}^{n} J_{j,n}^2\left(g\right)\right), \tag{2.11}$$

The formula above is exactly the likelihood process for a experiment given by observations

$$y_j = \frac{1}{\sqrt{2}} J_{j,n}\left(g\right) + \eta_j, \eta_j \text{ i.i.d.} N\left(0,1\right), j = 1, \ldots, n.$$

which is exactly equivalent to $E_{2,n}^1$

$$y_j = \frac{1}{\sqrt{2}} J_{j,n}\left(\log f\right) + \eta_j, \eta_j \text{ i.i.d.} N\left(0,1\right), j = 1, \ldots, n.$$

We may assume ζ_j and η_j reside on the common probability space with an appropriate coupling given in the next section. With a little abuse of notations we don't distinguish $\Lambda_{1,n}$ and $\Lambda_{2,n}$ from their versions. The Le Cam deficiency distance between $E_{1,n}^2$ and $E_{2,n}^1$ satisfies

$$\Delta\left(E_{1,n}^2, E_{2,n}^1\right) \le \frac{1}{2} \sup_{f \in \Sigma_n(f_0)} E\left|\Lambda_{1,n} - \Lambda_{2,n}\right|.$$

See lemma 6 in chapter 2 of Le Cam and Yang (2000) or proposition 2.2 in Nussbaum (1996). By a simple fact that the square of total variation distance is bounded by the Hellinger distance, we have

$$\Delta^2\left(E_{1,n}^2, E_{2,n}^1\right) \le \sup_{f \in \Sigma_n(f_0)} H^2\left(\Lambda_{1,n}, \Lambda_{2,n}\right).$$

Then if we can prove $H^2\left(\Lambda_{1,n}, \Lambda_{2,n}\right)$ converges to 0 uniformly over $\Sigma_n\left(f_0\right)$, we have local asymptotic equivalence between $E_{1,n}^2$ and $E_{2,n}^1$. So our goal now is to show

$$\sup_{f \in \Sigma_n(f_0)} H^2\left(\Lambda_{1,n}, \Lambda_{2,n}\right) \to 0.$$

Consider a partition of $[0,1]$ into subintervals

$$D_i = \left[(i-1)/2^{k_0}, i/2^{k_0}\right), \ i = 1, 2, \ldots, 2^{k_0} = \sqrt{n},$$

such that we may decompose the local experiments $E_{1,n}^2$ and $E_{2,n}^1$ into products of independent experiments. Similar technique has been used in Nussbaum (1996) (see also Grama and Nussbaum (1998, 2003)). Let $m_n = n/2^{k_0}$; in our setting this is the number of observations on the interval D_i. Recall

$$\Lambda_{1,n} = \exp\left(\frac{1}{\sqrt{2}}\sum_{j=1}^{n} \zeta_j\left(J_{j,n}(g)\right) - \frac{1}{2}\sum_{j=1}^{n}\left(J_{j,n}(g) - \log\left(1 + J_{j,n}(g)\right)\right)\right)$$

where $g = \log(f_0/f)$ and $\zeta_j = -\left(\xi_j^2 - 1\right)/\sqrt{2}$. Let I_i be the index of all observations $z(j)$ on the interval D_i. For each interval D_i we form a local likelihood process

$$\Lambda_{1,i,n} = \exp\left(\frac{1}{\sqrt{2}}\sum_{j \in I_i} \zeta_j J_{j,n}(g) - \frac{1}{2}\sum_{j \in I_i}\left(J_{j,n}(g) - \log\left(1 + J_{j,n}(g)\right)\right)\right).$$

Let $\Lambda_{2,i,n}$ be the corresponding local likelihood process for he likelihood process $\Lambda_{2,n}$ on the interval D_i. From the lemma 2.9 in appendix we have

$$H^2\left(\Lambda_{1,n}, \Lambda_{2,n}\right) \le \sum_{i=1}^{2^{k_0}} H^2\left(\Lambda_{1,i,n}, \Lambda_{2,i,n}\right).$$

We will derive nice upper bounds for $H^2(\Lambda_{1,i,n}, \Lambda_{2,i,n})$ such that the right side in the equation above goes to 0 uniformly. Let

$$\begin{aligned} S_i &= \exp\left(\frac{1}{\sqrt{2}}\sum_{j\in I_i}(\zeta_j - \eta_j) J_{j,n}(g)\right) \\ R_i &= \exp\left(-\frac{1}{2}\sum_{j\in I_i}\left(J_{j,n}(g) - \log(1 + J_{j,n}(g)) - \frac{1}{2}J_{j,n}^2(g)\right)\right), \end{aligned}$$

such that $\Lambda_{1,i,n}^{1/2}/\Lambda_{2,i,n}^{1/2} = S_i R_i$, then

$$\begin{aligned} \sum_{i=1}^{2^{k_0}} H^2(\Lambda_{1,i,n}, \Lambda_{2,i,n}) &= \sum_{i=1}^{2^{k_0}} E\left(\Lambda_{1,i,n}^{1/2}/\Lambda_{2,i,n}^{1/2} - 1\right)^2 \Lambda_{2,i,n} \\ &= \sum_{i=1}^{2^{k_0}} E(S_i R_i - R_i + R_i - 1)^2 \Lambda_{2,i,n} \le N_1 + N_2, \end{aligned}$$

where

$$N_1 = 2\sum_{i=1}^{2^{k_0}} (R_i - 1)^2, \; N_2 = 2\sum_{i=1}^{2^{k_0}} E(S_i - 1)^2 R_i^2 \Lambda_{2,i,n}. \tag{2.12}$$

Thus it is enough to show N_1 and N_2 converges to 0 uniformly over $\Sigma_n(f_0)$ so that $E_{1,n}^2$ and $E_{2,n}^1$ are locally asymptotically equivalent.

In the whole proof the symbols C and C_k denote generic positive constants.

Now we show N_1 converges to 0 uniformly over $\Sigma_n(f_0)$. The proof for N_2 is much more complicate and will be given later. Using Cauchy-Shwartz inequality and Jensen's inequality, and noting that $|g| = |\log f_0 - \log f| \le C|f_0 - f|$ by the mean value theorem, we have

$$\sum_{i=1}^{2^{k_0}}\left(\sum_{j\in I_i} J_{j,n}^3(g)\right)^2 \le \sqrt{n}\sum_{j=1}^{n} J_{j,n}^6(g) \le n^{3/2}\sum_{j=1}^{n} J_{j,n}(g^6) \le Cn^{3/2}\int |f - f_0|^6 \tag{2.13}$$

which converges to 0 uniformly over $\Sigma_n(f_0)$ from the equation (2.8) with $p = 6$. A three term Taylor expansion, combining with $|J_{j,n}(g)| \to 0$ implied by the equation above, yields

$$\left|J_{j,n}(g) - \log(1 + J_{j,n}(g)) - \frac{1}{2}J_{j,n}^2(g)\right| \le C\left|J_{j,n}^3(g)\right|. \tag{2.14}$$

Combining the equations (2.14) and (2.13), we have

$$\sum_{i=1}^{2^{k_0}} (\log R_i)^2 \leq Cn^{3/2} \int |f - f_0|^6 \to 0$$

uniformly over $\Sigma_n (f_0)$. Note that $|R_i - 1| = |\exp(\log R_i) - 1| \leq C |\log R_i|$, since $\log R_i$ are uniformly bounded which is implied in the equation above. So we have

$$N_1 = 2 \sum_{i=1}^{2^{k_0}} (R_i - 1)^2 \leq C_1 \sum_{i=1}^{2^{k_0}} (\log R_i)^2 \leq C_2 n^{3/2} \int |f - f_0|^6 \to 0 \tag{2.15}$$

uniformly over $\Sigma_n (f_0)$.

Thus the key step of establishing the local asymptotic equivalence is to prove N_2 converges to 0 uniformly over $\Sigma_n (f_0)$. Since all R_i are uniformly bounded, then we have

$$N_2 = 2 \sum_{i=1}^{2^{k_0}} E (S_i - 1)^2 R_i^2 \Lambda_{2,i,n} \leq C \sum_{i=1}^{2^{k_0}} E \int (S_i - 1)^2 \Lambda_{2,i,n}. \tag{2.16}$$

Before proving this result, we rewrite S_i in a convenient form by the Harr basis expansion such that we may apply the "good" coupling method proposed in the next section. Recall that

$$\left\{ \phi_{k_0 l} (x) , \; l = 1, ..., 2^{k_0}, \; \psi_{kl} (x) , \; l = 1, ..., 2^k \right\}$$

is the orthonormal Haar basis. Let

$$g_{kl} = \langle g, \phi_{kl} \rangle , \; \widetilde{g}_{kl} = \langle g, \psi_{kl} \rangle$$

For a sequence $(a_j)_{1 \leq j \leq n}$, we define the discrete Harr transform

$$a_{(k_0 l)} = \sum_{j=\frac{n}{2^{k_0}}(l-1)+1}^{\frac{n}{2^{k_0}} l} a_j, \; \widetilde{a}_{(kl)} = a_{(k+1,2l-1)} - a_{(k+1,2l)}$$

Then

$$\sum_{j=1}^{n} a_j J_{j,n}(g) = \sum_{l=1}^{2^{k_0}} 2^{k_0/2} a_{(k_0 l)} g_{k_0 l} + \sum_{k=k_0}^{k_1 - 1} \sum_{l=1}^{2^k} 2^{i/2} \widetilde{a}_{(kl)} \; \widetilde{g}_{kl},$$

so we can write

$$\begin{aligned} S_i &= \exp\left(\sum_{j\in I_i} \frac{1}{\sqrt{2}} \left(\zeta_j - \eta_j\right) J_{j,n}(g)\right) \\ &= \exp\left(\begin{array}{c} \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right) \\ + \frac{1}{\sqrt{2}} \sum_{k=k_0}^{k_1-1} \sum_{l\in I_k^i} 2^{k/2}\, \widetilde{g}_{kl} \left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)}\right) \end{array}\right) \end{aligned} \tag{2.17}$$

where $I_k^i = \left\{l : (i-1)/2^{k_0} < l/2^k \leq i/2^{k_0}\right\}$. From (2.16) we want to prove

$$\sum_{i=1}^{2^{k_0}} E\left(S_i - 1\right)^2 \Lambda_{2,i,n} \to 0 \tag{2.18}$$

uniformly over $\Sigma_n\left(f_0\right)$ such that N_2 converges to 0 uniformly. The proof will be given in section 2.4.

2.3 Construction of the Likelihoods on the Same Probability Space

We need to construct versions of $\zeta_{(k_0 i)}$, $\eta_{(k_0 i)}$, $\widetilde{\zeta}_{(kl)}$ and $\widetilde{\eta}_{(kl)}$ on the same probability space such that the equation (2.17) in the previous section is valid. Actually close matchings are necessary for pairs $\left(\zeta_{(k_0 i)}, \eta_{(k_0 i)}\right)$ and $\left(\widetilde{\zeta}_{(kl)}, \widetilde{\eta}_{(kl)}\right)$ such that S_i is around 1 which is required in (2.18). We use a quantile coupling approach, which was first introduced in the KMT paper, to construct a common probability space for those random variables. In this section we talk about upper bounds of $\zeta_{(k_0 i)} - \eta_{(k_0 i)}$ and $\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)}$ as the preparation of the proofs in the next section, section 2.4.

The following result motivates our coupling procedure, which leads to a different the approach from the KMT paper.

Proposition 2.1 *Suppose r.v.'s X, Y are independent, both with laws χ_n^2. Then the conditional law of X given $X+Y$ is the law of $(X+Y)B$, where B is a Beta*

$(n/2, n/2)$ *random variable:*

$$\mathcal{L}(X|X+Y) = \mathcal{L}((X+Y)B).$$

To define the coupling, let

$$\left\{X_l, l = 1, \ldots, 2^{k_0}, B_{kl}, l = 1, \ldots, 2^k, k = k_0, \ldots, k_1 - 1, k_1 = \log_2 n\right\}$$

be an array of independent r.v.'s such that X_l has a χ^2_m distribution for $m = 2^{k_1-k_0}$ and B_{kl} has a Beta distribution $B(2^{k_1-k-2}, 2^{k_1-k-2})$. Define

$$X_{k_0 l} = X_l, \ l = 1, \ldots, 2^{k_0}, \tag{2.19}$$

and recursively for $k = k_0, \ldots, k_1 - 1$

$$X_{k+1,2l-1} = X_{k,l}B_{kl}, \ X_{k+1,2l} = X_{k,l}(1 - B_{kl}), \ l = 1, \ldots, 2^k.$$

Lemma 2.1 *For every $k = k_0, \ldots, k_1$, the joint distribution of $X_{k,l}, \ l = 1, \ldots, 2^k$ is that of independent r.v.'s having a χ^2 distribution with 2^{k_1-k} degrees of freedom.*

Proof: For $k = k_0$ the claim follows from (2.19). Assume the claim is proved for k; then the pairs $(X_{k+1,2l-1}, X_{k+1,2l})$ are independent. Furthermore, $X_{k,l} = X_{k+1,2l-1} + X_{k+1,2l}$ and $X_{k,l}$ has a χ^2_{2m} distribution with $m = 2^{k_1-(k+1)}$. It follows

$$\mathcal{L}(X_{k+1,2l-1}|X_{k,l}) = X_{k,l}B_{kl}, \ \mathcal{L}(X_{k+1,2l}|X_{k,l}) = X_{k,l}(1 - B_{kl})$$

and B_{il} has a Beta $B(m/2, m/2)$ distribution. From the proposition 2.1, it follows that the conditional laws of $X_{k+1,2l-1}$ and $X_{k+1,2l}$ given their sum are exactly the conditional laws of two independent χ^2_m variables given their sum. Thus $X_{k+1,2l-1}$ and $X_{k+1,2l}$ are independent χ^2_m variables.

Our basic probability space is the one of the array

$$\left\{X_l, l = 1, \ldots, 2^{k_0}, B_{kl}, l = 1, \ldots, 2^k, k = k_0, \ldots, k_1 - 1\right\}$$

defined above. We construct ζ_k, $k = 1, \ldots, n$ by

$$\zeta_l = -\left(X_{k_1,j} - 1\right)/\sqrt{2},\ l = 1, \ldots, 2^{k_1}.$$

Then

$$\begin{aligned} \zeta_{(k_0 l)} &= -\left(X_l - 2^{k_1-k_0}\right)/\sqrt{2},\ l = 1, \ldots, 2^{k_0}, \\ \zeta_{(kl)} &= -\left(X_{kl} - 2^{k_1-k}\right)/\sqrt{2},\ l = 1, \ldots, 2^{k}, k = k_0, \ldots, k_1, \\ \tilde{\zeta}_{(kl)} &= \left(X_{k+1,2l} - X_{k+1,2l-1}\right)/\sqrt{2} \\ &= \sqrt{2}X_{k,l}\left(1/2 - B_{kl}\right),\ l = 1, \ldots, 2^{k}, k = k_0, \ldots, k_1 - 1 \end{aligned} \tag{2.20}$$

We construct $\eta_{(k_0 l)}$ and $\tilde{\eta}_{(kl)}$ as follows. Let $G_{k_1-k_0}$ the distribution function of $\zeta_{(k_0 1)}/2^{(k_1-k_0)/2}$ and Φ be the standard normal d.f.; we set

$$\eta_{(k_0 l)} = 2^{(k_1-k_0)/2}\Phi^{-1} \circ G_m\left(\zeta_{(k_0 l)}/2^{(k_1-k_0)/2}\right) \tag{2.21}$$

then $\eta_{(k_0 l)}$ are independent $N(0, 2^{k_1-k_0})$ random variables. Furthermore, for $k = k_0, \ldots, k_1 - 1$ let $\tilde{G}_{k_1-(k+1)}$ be the distribution function of $2^{(k_1-k+1)/2}\left(1/2 - B_{k1}\right)$; from the equation (2.45) we set

$$\tilde{\eta}_{(kl)} = 2^{(k_1-k)/2}\Phi^{-1} \circ \tilde{G}_{k_1-(k+1)}\left(2^{(k_1-k+1)/2}\left(1/2 - B_{kl}\right)\right), k = k_0, \ldots, k_1 - 1 \tag{2.22}$$

then $\tilde{\eta}_{(kl)}$ are independent $N(0, 2^{k_1-k})$ random variables. Moreover, in the array

$$\eta_{(k_0 l)}, l = 1, \ldots, 2^{k_0}, \tilde{\eta}_{(kl)},\ l = 1, \ldots, 2^{k}, k = k_0, \ldots, k_1 - 1$$

all r.v.'s are independent.

Remark 2.3 *Above we used direct quantile coupling, using the respective d.f.'s of random variables. However we may also use other "good" couplings of chi-squares and beta's with normals, it is important however that $\eta_{(k_0 l)}$ is constructed from $\zeta_{(k_0 l)}$ and $\tilde{\eta}_{(kl)}$ from B_{kl}, to guarantee the independence properties.*

In the section "Quantile coupling for Beta distributions", we prove the following result:

Let Z be a standard normal random variable. For every n, there is a mapping $T_n : R \mapsto R$ such that the random variable $B_n = T_n(Z)$ has the Beta $(n/2, n/2)$ law and

$$\left|n\left(1/2 - B_n\right) - \left(n^{1/2}/2\right) Z\right| \le \frac{C}{\sqrt{n}}\left(1 + |Z|^3\right), \text{ when } |Z| \le \varepsilon\sqrt{n},$$

where $C, \varepsilon > 0$ do not depend on n.

This result implies

$$\begin{aligned} \tilde{\zeta}_{(kl)} - \tilde{\eta}_{(kl)} &= X_{k,l}\left(1 - 2B_{kl}\right)/\sqrt{2} - \frac{X_{k,l}}{n/2^k}\tilde{\eta}_{(kl)} + \frac{X_{k,l}}{n/2^k}\tilde{\eta}_{(kl)} - \tilde{\eta}_{(kl)} \\ &= \frac{X_{k,l}}{n/2^k}\left(\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right) + \left(\frac{X_{k,l}}{n/2^k} - 1\right)\tilde{\eta}_{(kl)}, \end{aligned}$$

where

$$\begin{aligned} &\left|\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right| \\ \le\ & C_1\left(\frac{1}{\sqrt{n/2^k}} + \frac{\left|\tilde{\eta}_{(kl)}\right|^3}{\left(n/2^k\right)^2}\right) \quad (2.23) \\ \le\ & C_2\left(\frac{1}{\sqrt{n/2^k}} + \frac{\left|\tilde{\eta}_{(kl)}\right|^2}{n/2^k}\right), \text{ when } \left|\tilde{\eta}_{(kl)}\right| \le \varepsilon\sqrt{n/2^k} \quad (2.24) \end{aligned}$$

for some C_1, C_2 and *$\varepsilon > 0$. We define*

$$\widetilde{G}_{kl} = \left\{\left|X_{k,l} - n/2^k\right| \le \left(n/2^k\right)^{2/3},\ \left|\tilde{\eta}_{(kl)}\right| \le \varepsilon\left(n/2^k\right)^{2/3}\right\},$$

then on $\widetilde{G}_{kl}$ obviously we have

$$\left|\tilde{\zeta}_{(kl)} - \tilde{\eta}_{(kl)}\right| \le C\left(1 + \frac{\left|\tilde{\eta}_{(kl)}\right|}{\left(n/2^k\right)^{1/3}} + \frac{\left|\tilde{\eta}_{(kl)}\right|^2}{n/2^k}\right) \quad (2.25)$$

for some $C > 0$.

The KMT result implies

$$\left|\zeta_{(k_0 l)} - \eta_{(k_0 l)}\right| \le C\left(1 + \frac{\zeta^2_{(k_0 l)}}{n/2^{k_0}}\right), \text{ when } \left|\zeta_{(k_0 l)}\right| \le \varepsilon n/2^{k_0}, \tag{2.26}$$

for some $C, \varepsilon > 0$ *(see Komlós, Major, and Tusnády (1976)). By similar arguments in the proof of corollary (2.1), we have*

$$\left|\zeta_{(k_0 l)} - \eta_{(k_0 l)}\right| \le C\left(1 + \frac{\eta^2_{(k_0 l)}}{n/2^{k_0}}\right), \text{ when } \left|\eta_{(k_0 l)}\right| \le \varepsilon n/2^{k_0} \tag{2.27}$$

on $G_{k_0 l}$ *with* $G_{k_0 l} = \left\{\left|\eta_{(k_0 l)}\right| \le \varepsilon n/2^{k_0}\right\}$, *for some* $C, \varepsilon > 0$.

2.4 Proof of the Main Result

In this section, we use the quantile coupling bounds given in the section 2.3 to show N_2 defined in (2.12) converges to 0 uniformly over $\Sigma_n(f_0)$, which leads to the asymptotic equivalence between $E_{1,n}$ and $E_{2,n}$ locally. Recall that it is enough to prove

$$\sum_{i=1}^{2^{k_0}} E\left(S_i - 1\right)^2 \Lambda_{2,i,n} \to 0$$

uniformly over $\Sigma_n(f_0)$, where

$$S_i = \exp\left(\frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right) + \frac{1}{\sqrt{2}} \sum_{k=k_0}^{k_1 - 1} \sum_{l \in I_k^i} 2^{k/2}\, \widetilde{g}_{kl}\left(\widetilde{\zeta}_{kl} - \widetilde{\eta}_{kl}\right)\right).$$

Define the filtration $\mathcal{F}_k$ as follows

$$\begin{aligned}
\mathcal{F}_{k_0-1} &= \{\emptyset, \Omega\}, \mathcal{F}_{k_0} = \sigma\left(X_l, l = 1, \ldots, 2^{k_0}\right), \\
\mathcal{F}_k &= \sigma\left(X_l, l = 1, \ldots, 2^{k_0}, B_{il}, j = 1, \ldots, 2^k, i = k_0, \ldots, k-1\right), \\
k &= k_0 + 1, \ldots, k_1 = \log_2 n.
\end{aligned}$$

such that we may define $S_i = U_i V_i$ with

$$U_i = \exp\begin{pmatrix} \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)}\right) + \\ \frac{1}{\sqrt{2}} \sum_{k=k_0}^{k_1-1} \sum_{l \in I_k^i} 2^{k/2}\, \widetilde{g}_{kl}\left(\widetilde{\zeta}_{kl} - \widetilde{\eta}_{kl} - \widetilde{\mu}_{kl}\right) \end{pmatrix},$$

and

$$V_i = \exp\left(\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)} + \frac{1}{\sqrt{2}}\sum_{k=k_0}^{k_1-1}\sum_{l\in I_k^i} 2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{kl}\right),$$

where

$$\begin{aligned}\mu_{(k_0 i)} &= E_{P_2}\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right)\\ \widetilde{\mu}_{(kl)} &= E_{P_2}\left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)}|\mathcal{F}_k\right),\end{aligned}$$

and

$$E_{P_2}(\cdot) \triangleq E(\cdot)\,\Lambda_{2,n} \tag{2.28}$$

.Then we have

$$\sum_{i=1}^{2^{k_0}} E\left(S_i - 1\right)^2 \Lambda_{2,i,n} = \sum_{i=1}^{2^{k_0}} E\left(U_i V_i - U_i + U_i - 1\right)^2 \Lambda_{2,i,n} \le 2N_{21} + 2N_{22},$$

where

$$N_{21} = \sum_{i=1}^{2^{k_0}} E_{P_2}\left(U_i - 1\right)^2,\ N_{22} = \sum_{i=1}^{2^{k_0}} E_{P_2}\left(V_i - 1\right)^2 U_i^2. \tag{2.29}$$

We will show that both N_{21} and N_{22} converge to 0 uniformly over $\Sigma_n(f_0)$, which implies N_2 converges to 0 uniformly over $\Sigma_n(f_0)$.

2.4.1 An Upper Bound of N_{21}

We show

$$N_{21} \le C\log n\left[\sum_{l=1}^{2^{k_0}} 2^{k_0} g_{k_0 l}^2 + \sum_{k=k_0}^{k_1-1}\sum_{l=1}^{2^k}\left(2^k\ \widetilde{g}_{kl}^2 + 2^{5k}\ \widetilde{g}_{kl}^6 + 2^k n^2 g_{kl}^6\right)\right] \tag{2.30}$$

which converges to zero uniformly over $\Sigma_n(f_0)$ from the lemma (2.13). The following lemma is the key result to establish the upper bound above.

Lemma 2.2 *Under the assumption For $f \in \Sigma_n(f_0)$, we have*

$$\begin{aligned}
0 &\le \log E_{P_2} \exp\left(K\frac{m}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)}\right)\right) \\
&\le CK^2 2^{k_0} g_{k_0 i}^2 \left(1 + n^2 g_{k_0 i}^4\right) \\
0 &\le \log E_{P_2} \exp\left(K\frac{m}{\sqrt{2}} 2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{kl} - \widetilde{\eta}_{kl} - \widetilde{\mu}_{kl}\right)\right) \\
&\le CK^2 2^k \widetilde{g}_{kl}^2 \left(1 + n\widetilde{g}_{kl}^2 + n g_{kl}^2 \cdot n\widetilde{g}_{kl}^2\right).
\end{aligned}$$

Now let's apply lemma 2.2 to establish the upper bound of N_{21} in (2.30). Note that Hölder inequality gives

$$\begin{aligned}
E_{P_2} U_i^m &\le \left(E_{P_2} \exp\left(K\frac{m}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)}\right)\right)\right)^{1/K} \\
&\quad \cdot \prod_{k=k_0}^{k_1-1} \left(\prod_{l \in I_k^i} E_{P_2}\left[\exp\left(K\frac{m}{\sqrt{2}} 2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{kl} - \widetilde{\eta}_{kl} - \widetilde{\mu}_{kl}\right)\right)\right]\right)^{1/K},
\end{aligned}$$

where $K = k_1 - k_0 + 1 \asymp \log n$, m is a finite positive number. Then from lemma 2.2 we have

$$0 \le \log E_{P_2} U_i^m \le CK\left[2^{k_0} g_{k_0 i}^2\left(1 + n^2 g_{k_0 i}^4\right) + \sum_{k=k_0}^{k_1-1}\sum_{l \in I_k^i} 2^k \widetilde{g}_{kl}^2\left(1 + n\widetilde{g}_{kl}^2 + n g_{kl}^2 \cdot n\widetilde{g}_{kl}^2\right)\right].$$

Note that $n = 2^{2k_0} \le 2^{2k}$ for $k \ge k_0$. Basic inequalities $ab \le (a^2 + b^2)/2$ and $abc \le (a^3 + b^3 + c^3)/3$ imply

$$\begin{aligned}
2^k \widetilde{g}_{kl}^2 \cdot n\widetilde{g}_{kl}^2 &\le 2^{3k}\ \widetilde{g}_{kl}^4 \le \frac{1}{2}\left(2^k\ \widetilde{g}_{kl}^2 + 2^{5k}\ \widetilde{g}_{kl}^6\right) \\
2^k \widetilde{g}_{kl}^2 \cdot n\widetilde{g}_{kl}^2 \cdot n g_{kl}^2 &\le 2^{5k/3}\ \widetilde{g}_{kl}^2 \cdot 2^{5k/3}\ \widetilde{g}_{kl}^2 \cdot \frac{n}{2^{k/3}} g_{kl}^2 \\
&\le \frac{2}{3} 2^{5k}\ \widetilde{g}_{kl}^6 + \frac{1}{3} 2^k n^2 g_{kl}^6
\end{aligned}$$

Thus we may write

$$0 \le \log E_{P_2} U_i^m \le CK\left[2^{k_0} g_{k_0 i}^2\left(1 + n^2 g_{k_0 i}^4\right) + \sum_{k=k_0}^{k_1-1}\sum_{l \in I_k^i}\left(2^k\ \widetilde{g}_{kl}^2 + 2^{5k}\ \widetilde{g}_{kl}^6 + 2^k n^2 g_{kl}^6\right)\right].$$

Lemma 2.13 implies that the right side of the equation above converges to 0 uniformly, then

$$0 \leq E_{P_2} U_i^m - 1 = \exp\left(\log E_{P_2} U_i^m\right) - 1 \leq C \log E_{P_2} U_i^m.$$

Combining two inequalities above we have

$$\begin{aligned}
\sum_{i=1}^{2^{k_0}} E_{P_2} \left(U_i - 1\right)^2 &= \sum_{i=1}^{2^{k_0}} E_{P_2} \left(U_i^2 - 1\right) + 2 \sum_{i=1}^{2^{k_0}} E_{P_2} \left(1 - U_i\right) \\
&\leq \sum_{i=1}^{2^{k_0}} E_{P_2} \left(U_i^2 - 1\right) \\
&\leq CK \left[\sum_{l=1}^{2^{k_0}} 2^{k_0} g_{k_0 l}^2 + \sum_{k=k_0}^{k_1 - 1} \sum_{l=1}^{2^k} \left(2^k\ \widetilde{g}_{kl}^2 + 2^{5k}\ \widetilde{g}_{kl}^6 + 2^k n^2 g_{kl}^6\right) \right].
\end{aligned}$$

The proof of the lemma (2.2) is based on some estimates stated and proved below.

Lemma 2.3 *Under the assumption For* $f \in \Sigma_n\left(f_0\right)$, *for a constant* $m > 0$ *there is a constant* $C > 0$ *such that*

$$\begin{aligned}
E_{P_2} \exp\left(m \left| K \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)}\right) \right| \right) &\leq C \\
E_{P_2} \exp\left(m \left| K \frac{1}{\sqrt{2}} 2^{k/2}\ \widetilde{g}_{kl} \left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)} - \widetilde{\mu}_{(kl)}\right) \right| \right) &\leq C.
\end{aligned}$$

Proof: (i)Since

$$\mu_{(k_0 i)} = E_{P_2} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right)$$

by the definition, Jensen's inequality implies

$$\begin{aligned}
\exp\left(m \left| \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \mu_{(k_0 i)} \right| \right) &\leq \exp\left(E_{P_2} m \left| M \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right) \right| \right) \\
&\leq E_{P_2} \exp\left(m \left| M \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right) \right| \right),
\end{aligned}$$

then we have

$$\begin{aligned}
& E_{P_2} \exp\left(m\left|M\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}-\mu_{(k_0i)}\right)\right|\right) \\
\leq\ & E_{P_2} \exp\left(2m\left|M\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}\right)\right|\right).
\end{aligned}$$

Set for brevity

$$G_{k_0i} = \left\{\left|\eta_{(k_0i)}\right| \leq \varepsilon\frac{n}{2^{k_0}}\right\},$$

where ε is chosen such that (2.27) holds, and $G^c_{k_0i}$ is the complement of G_{k_0i}. We write

$$E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}\right)\right|\right) = Q_1 + Q_2$$

where

$$\begin{aligned}
Q_1 &= E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}\right)\right|\right)\mathbf{1}\left(G_{k_0i}\right) \\
Q_2 &= E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}\right)\right|\right)\mathbf{1}\left(G^c_{k_0i}\right).
\end{aligned}$$

The equation (2.27) implies

$$Q_1 \leq E_{P_2} \exp\left(2Cm\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(1+\frac{\eta^2_{(k_0i)}}{n/2^{k_0}}\right)\right|\right)$$

Since $\eta_{(k_0i)} \sim N\left(\sqrt{\frac{n}{2}}g_{k_0i}, n/2^{k_0}\right)$ under the probability measure P_2 defined in (2.28), we have

$$\begin{aligned}
Q_1 &\leq E\exp\left(2Cm\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(1+\frac{2\eta^2_{(k_0i)}}{n/2^{k_0}}+2\frac{n}{2}g^2_{k_0i}\right)\right|\right) \\
&= \exp\left(\sqrt{2}CmK\left(\left(2^{k_0}g^2_{k_0i}\right)^{1/2}+\left(2^{k_0}n^2g^6_{k_0i}\right)^{1/2}\right)\right) \\
&\quad \cdot E\exp\left(2\sqrt{2}CmK\left|2^{k_0/2}g_{k_0i}\right|\frac{\eta^2_{(k_0i)}}{n/2^{k_0}}\right),
\end{aligned}$$

then from lemma 2.13 it is easy to see Q_1 is bounded.

We see

$$\begin{aligned} Q_2 &= E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}\right)\right|\right)\mathbf{1}\left(G^c_{k_0i}\right) \\ &= E\exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\left(\zeta_{(k_0i)}-\eta_{(k_0i)}\right)\right|\right) \\ &\quad \cdot\exp\left(\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\eta_{(k_0i)}-\frac{n}{4}g^2_{k_0i}\right)\mathbf{1}\left(G^c_{k_0i}\right) \\ &\leq E\exp\left(2mK\left|\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\right|\left|\zeta_{(k_0i)}\right|\right) \\ &\quad \cdot\exp\left((2mK+1)\left|\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\right|\left|\eta_{(k_0i)}\right|\right)\mathbf{1}\left(G^c_{k_0i}\right) \end{aligned}$$

Hölder inequality implies

$$\begin{aligned} Q_2 &\leq \left[E\exp\left(6mK\left|\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\right|\left|\zeta_{(k_0i)}\right|\right)\right]^{1/3} \\ &\quad\cdot\left[E\exp\left(3\left(2mK+1\right)\left|\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\right|\left|\eta_{(k_0i)}\right|\right)\right]^{1/3}\left[E\mathbf{1}\left(G^c_{k_0i}\right)\right]^{1/3} \end{aligned}$$

¿From lemma 2.16 we have

$$\begin{aligned} E\exp\left(6mK\left|\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\right|\left|\zeta_{(k_0i)}\right|\right) &\leq C_1+\exp\left(C_2K^2ng^2_{k_0i}\right) \\ E\exp\left(3\left(2mK+1\right)\left|\frac{1}{\sqrt{2}}2^{k_0/2}g_{k_0i}\right|\left|\eta_{(k_0i)}\right|\right) &\leq C_1+\exp\left(C_2K^2ng^2_{k_0i}\right) \\ \left[E\mathbf{1}\left(G^c_{k_0i}\right)\right]^{1/2} &\leq C_1\exp\left(-C_2\frac{n}{2^{k_0}}\right) \end{aligned}$$

and from lemma 2.13 we have

$$K^2ng^2_{k_0i}\ll\frac{n}{2^{k_0}},$$

so Q_2 is also bounded. Since Q_1 and Q_2 are bounded, the first inequality is established.

(ii) The proof of the second inequality is similar to that of the first one. The proof given below will just emphasize parts which are different from the proof of the first inequality. We have

$$
\begin{aligned}
& E_{P_2} \exp\left(m\left|K\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{(kl)}-\widetilde{\eta}_{(kl)}-\widetilde{\mu}_{(kl)}\right)\right|\right) \\
\leq\ & E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{(kl)}-\widetilde{\eta}_{(kl)}\right)\right|\right) = M_1 + M_2
\end{aligned}
$$

where

$$
\begin{aligned}
M_1 &= E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{(kl)}-\widetilde{\eta}_{(kl)}\right)\right|\right) 1\left(G_{kl}\right) \\
M_2 &= E_{P_2} \exp\left(2m\left|K\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{(kl)}-\widetilde{\eta}_{(kl)}\right)\right|\right) 1\left(G_{kl}^c\right).
\end{aligned}
$$

On $\widetilde{G}_{kl} = \left\{\left|X_{k,l}-n/2^k\right| \leq \left(n/2^k\right)^{2/3},\ \left|\tilde{\eta}_{(kl)}\right| \leq \varepsilon\left(n/2^k\right)^{2/3}\right\}$, from the equation (2.25). and noting that $\eta_{(k_0 i)} \sim N\left(\sqrt{\frac{n}{2}}\widetilde{g}_{kl}, n/2^k\right)$ under P_2, we have

$$
\begin{aligned}
M_1 &= E_{P_2} \exp\left(C2mK\left|\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right|\left(1+\frac{\left|\tilde{\eta}_{(kl)}\right|}{\left(n/2^k\right)^{1/3}}+\frac{\left|\tilde{\eta}_{(kl)}\right|^2}{n/2^k}\right)\right) 1\left(G_{kl}\right) \\
&\leq \exp\left(C2mK\left|\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right|\left(1+\left(n/2^k\right)^{1/6}\sqrt{n}\left|\widetilde{g}_{kl}\right|+n\widetilde{g}_{kl}^2\right)\right) \\
&\quad \cdot E\exp\left(C2m\left|\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right|\left(\left(n/2^k\right)^{1/6}\frac{\left|\tilde{\eta}_{(kl)}\right|}{\left(n/2^k\right)^{1/2}}+\frac{\left|\tilde{\eta}_{(kl)}\right|^2}{n/2^k}\right)\right) \\
&\leq C_1\left(1+K\left(2^k\left|\ \widetilde{g}_{kl}\right|^2\right)^{1/2}+K\left(2^k n^2\widetilde{g}_{kl}^6\right)^{1/3}+K\left(2^k n^2\ \left|\widetilde{g}_{kl}\right|^6\right)^{1/2}\right)
\end{aligned}
$$

which is bounded from lemma 2.13.

And on $\widetilde{G}_{kl}^c$ similarly we have

$$
\begin{aligned}
M_2 &\leq E1\left(\widetilde{G}_{kl}^c\right) E\exp\left(C\left|K\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\zeta}_{(kl)}\right|\right)^{1/3} \\
&\quad \cdot E\exp\left(C\left|K\frac{1}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\eta}_{(kl)}\right|\right)^{1/3} \\
&\leq C_1\exp\left(-C_2\left(\frac{n}{2^k}\right)^{1/3}\right)\left(C_1+\exp\left(C_2K^2n\widetilde{g}_{kl}^2\right)\right)^2
\end{aligned}
$$

which is bounded due to the fact that $K^2 n\widetilde{g}_{ki}^2 \ll \left(\frac{n}{2^k}\right)^{1/3}$, i.e., $K^2 2^k n^2 \widetilde{g}_{ki}^6 \ll 1$.

Remark 2.4 *Actuarially the proof above shows that for a constant $m > 0$ there is a constant $C > 0$ such that*

$$
\begin{aligned}
& E_{P_2} \exp\left(m \left| K \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \mu_{(k_0 i)} \right| \right) \\
\leq \; & E_{P_2} \exp\left(m \left| K \frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} \right) \right| \right) \leq C \\
& E_{P_2} \exp\left(m \left| K \frac{1}{\sqrt{2}} 2^{k/2} \; \widetilde{g}_{kl} \widetilde{\mu}_{(kl)} \right| \right) \\
\leq \; & E_{P_2} \exp\left(m \left| K \frac{1}{\sqrt{2}} 2^{k/2} \; \widetilde{g}_{kl} \left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)} \right) \right| \right) \leq C.
\end{aligned}
$$

Proof of lemma 2.2: (i) From lemma 2.14 in appendix and lemma 2.3 we have

$$
\begin{aligned}
0 \;\leq\; & \log E_{P_2} \exp\left(K \frac{m}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)} \right) \right) \\
\leq\; & C K^2 2^{k_0} g_{k_0 i}^2 \left[E_{P_2} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)} \right)^4 \right]^{1/2}
\end{aligned}
$$

and Jensen's inequality implies

$$
\mu_{(k_0 l)}^4 \leq E \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} \right)^4
$$

then

$$
\begin{aligned}
0 \;\leq\; & \log E_{P_2} \exp\left(K \frac{m}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)} \right) \right) \\
\leq\; & C K^2 2^{k_0} g_{k_0 l}^2 \left[E_{P_2} \left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} \right)^4 \right]^{1/2}.
\end{aligned}
$$

¿From (2.27), we may write

$$
\begin{aligned}
\left| \zeta_{(k_0 i)} - \eta_{(k_0 i)} \right| \;=\; & \left| \zeta_{(k_0 i)} - \eta_{(k_0 i)} \right| \left(\mathbf{1}\left(G_{k_0 i} \right) + \mathbf{1}\left(G_{k_0 i}^c \right) \right) \\
\leq\; & C \left(1 + \frac{\eta_{(k_0 i)}^2}{n / 2^{k_0}} \right) \mathbf{1}\left(G_{k_0 i} \right) + \left| \zeta_{(k_0 i)} - \eta_{(k_0 i)} \right| \mathbf{1}\left(G_{k_0 i}^c \right).
\end{aligned}
$$

By similar arguments in the lemma 2.3 we have

$$
\begin{aligned}
& E_{P_2} \left| \zeta_{(k_0 i)} - \eta_{(k_0 i)} \right| \mathbf{1}\left(G_{k_0 i}^c \right) \\
\leq\; & \left(E \left| \zeta_{(k_0 i)} - \eta_{(k_0 i)} \right|^3 \right)^{1/3} \left(E \exp\left(\frac{3}{\sqrt{2}} 2^{k_0/2} g_{k_0 i} \eta_{(k_0 i)} \right) \right)^{1/3} \left(E \mathbf{1}\left(G_{k_0 i}^c \right) \right)^{1/3} \leq C.
\end{aligned}
$$

Note that

$$E_{P_2}\left(\frac{\eta^2_{(k_0 i)}}{n/2^{k_0}}\right) = 1 + \frac{1}{2} n g^2_{k_0 i},$$

then

$$0 \leq \log E_{P_2} \exp\left(K \frac{m}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)} - \mu_{(k_0 i)}\right)\right) \leq C K^2 2^{k_0} g^2_{k_0 i}\left(1 + n^2 g^4_{k_0 l}\right).$$

(ii) Similarly we can show

$$\begin{aligned}
0 &\leq \log E_{P_2} \exp\left(K \frac{m}{\sqrt{2}} 2^{k/2}\ \widetilde{g}_{kl}\left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)} - \widetilde{\mu}_{(kl)}\right)\right) \\
&\leq C_1 K^2 2^k \widetilde{g}^2_{kl}\left[E_{P_2}\left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)}\right)^4\right]^{1/2} \\
&\leq C_2 K^2 2^k \widetilde{g}^2_{kl}\left[\begin{array}{c} E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\left(\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right) + \left(\frac{X_{k,l}}{n/2^k} - 1\right)\tilde{\eta}_{(kl)}\right)^4 \\ \cdot\left(1\left(\widetilde{G}_{kl}\right) + 1\left(\widetilde{G}^c_{kl}\right)\right)\end{array}\right]^{1/2} \\
&\leq C_3 K^2 2^k \widetilde{g}^2_{kl}\left(1 + n\widetilde{g}^2_{kl} + n g^2_{kl} \cdot n\widetilde{g}^2_{kl}\right).
\end{aligned}$$

2.4.2 An upper bound of N_{22}

Now we show

$$N_{22} \leq C \log n\left[\sum_{l=1}^{2^{k_0}} 2^{k_0} g^2_{k_0 l} + \sum_{k=k_0}^{k_1 - 1}\sum_{l=1}^{2^k}\left(2^k\ \widetilde{g}^2_{kl} + 2^{5k}\ \widetilde{g}^6_{kl} + 2^k n^2 g^6_{kl}\right)\right] \tag{2.31}$$

which converges to zero uniformly over $\Sigma_n(f_0)$ from the lemma (2.13). Recall that

$$V_i = \exp\left(\frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\mu_{(k_0 i)} + \frac{1}{\sqrt{2}}\sum_{k=k_0}^{k_1-1}\sum_{l\in I^i_k} 2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{kl}\right).$$

Since $|x-y| \leq (x+y)\left|\log(x/y)\right|$ for all positive x and y,

$$E\left(V_i - 1\right)^2 U_i^2 \Lambda_{2,i,n} \leq E_{P_2}\left[\begin{array}{c}\left(\frac{1}{\sqrt{2}} 2^{k_0/2} g_{k_0 i}\mu_{(k_0 i)} + \frac{1}{\sqrt{2}}\sum_{k=k_0}^{k_1-1}\sum_{l\in I^i_k} 2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)}\right)^2 \\ \cdot\left(V_i + 1\right)^2 U_i^2 \Lambda_{2,i,n}\end{array}\right]$$

so that by applying Cauchy-Schwartz inequality three times we obtain,

$$\begin{aligned}
& E_{P_2}\left(V_i-1\right)^2 U_i^2 \Lambda_{2,i,n} \\
\leq\ & K E_{P_2}\left(2^{k_0} g_{k_0 i}^2 \mu_{(k_0 i)}^2+\sum_{k=k_0}^{k_1-1}\left(\sum_{l \in I_k^i} 2^{k/2}\ \widetilde{g}_{kl} \widetilde{\mu}_{(kl)}\right)^2\right)\left(V_i+1\right)^2 U_i^2 \\
\leq\ & \log n \cdot E_{P_2}\left(2^{k_0} g_{k_0 i}^2 \mu_{(k_0 i)}^2+\sum_{k=k_0}^{k_1-1} \sum_{l \in I_k^i} \frac{2^k}{2^{k_0}} 2^k\ \widetilde{g}_{kl}^2 \widetilde{\mu}_{(kl)}^2\right)\left(V_i+1\right)^2 U_i^2 \\
\leq\ & \log n \cdot\left(2^{k_0} g_{k_0 i}^2 \mu_{(k_0 i)}^2+\sum_{k=k_0}^{k_1-1} \sum_{l \in I_k^i} \frac{2^{2k}}{2^{k_0}}\ \widetilde{g}_{kl}^2\left(E_{P_2} \widetilde{\mu}_{(kl)}^4\right)^{1/2}\right)\left(E_{P_2}\left(V_i+1\right)^4 U_i^4\right)^{1/2} .
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\mu}_{(kl)} &= E_{P_2}\left(\widetilde{\zeta}_{(kl)}-\widetilde{\eta}_{(kl)} | \mathcal{F}_k\right) \\
&= \frac{X_{k,l}}{n/2^k} E_{P_2}\left(\frac{n}{2^k}\left(1-2 B_{kl}\right)/\sqrt{2}-\tilde{\eta}_{(kl)}\right)+\left(\frac{X_{k,l}}{n/2^k}-1\right) E_{P_2} \tilde{\eta}_{(kl)} .
\end{aligned}$$

We will show

Lemma 2.4 *For $f \in \Sigma_n\left(f_0\right)$, we have*

$$\begin{aligned}
\mu_{(k_0 i)}^2 &\leq C\left(1+n^2 g_{k_0 i}^4\right) \\
\left(E \widetilde{\mu}_{(kl)}^4 \Lambda_{2,i,n}\right)^{1/2} &\leq n \widetilde{g}_{kl}^2\left(1+n g_{kl}^2+n \widetilde{g}_{kl}^2\right) \\
E\left(V_i+1\right)^4 U_i^4 \Lambda_{2,i,n} &\leq C .
\end{aligned}$$

Now we apply lemma 2.4 to establish (2.31). Basic inequalities $ab \leq \left(a^2+b^2\right)/2$ and $abc \leq \left(a^3+b^3+c^3\right)/3$ for $a+b+c \geq 0$ imply

$$\begin{aligned}
\frac{2^{2k}}{2^{k_0}}\ \widetilde{g}_{kl}^2 n \widetilde{g}_{kl}^2 &\leq 2^{k/2}\ \widetilde{g}_{kl} 2^{5/2k} \widetilde{g}_{kl}^3 \leq \frac{1}{2}\left(2^k\ \widetilde{g}_{kl}^2+2^{5k}\ \widetilde{g}_{kl}^6\right) \\
\frac{2^{2k}}{2^{k_0}}\ \widetilde{g}_{kl}^2 \cdot n \widetilde{g}_{kl}^2 \cdot n g_{kl}^2 &\leq 2^{5k/3}\ \widetilde{g}_{kl}^2 \cdot 2^{5k/3}\ \widetilde{g}_{kl}^2 \cdot \frac{n}{2^{k/3}} g_{kl}^2 \leq \frac{2}{3} 2^{5k}\ \widetilde{g}_{kl}^5+\frac{1}{3} 2^k n^2 g_{kl}^6,
\end{aligned}$$

then we have

$$N_{22} \leq C_1 \log n \sum_i \begin{pmatrix} 2^{k_0} g_{k_0 i}^2 \left(1 + n^2 g_{k_0 i}^4\right) + \\ \sum_{k=k_0}^{k_1-1} \sum_{l \in I_k^i} \frac{2^{2k}}{2^{k_0}} \ \widetilde{g}_{kl}^2 n \widetilde{g}_{kl}^2 \left(1 + n g_{kl}^2 + n \widetilde{g}_{kl}^2\right) \end{pmatrix} \quad (2.32)$$
$$\leq C_2 \log n \left[\sum_i 2^{k_0} g_{k_0 i}^2 + \sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} \left(2^k \ \widetilde{g}_{kl}^2 + 2^{5k} \ \widetilde{g}_{kl}^6 + 2^k n^2 g_{kl}^6\right) \right].$$

Proof of lemma 2.4:

(i) From the equation (2.27) we have

$$\begin{aligned} \mu_{(k_0 i)}^2 &\leq E\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right)^2 \Lambda_{2,i,n} \\ &\leq CE\left(1 + \frac{\eta_{(k_0 i)}^2}{n/2^{k_0}}\right)^2 \mathbf{1}\left(G_{k_0 i}\right) \Lambda_{2,i,n} + E\left(\zeta_{(k_0 i)} - \eta_{(k_0 i)}\right)^2 \mathbf{1}\left(G_{k_0 i}^c\right) \Lambda_{2,i,n} \\ &\leq C\left(1 + n^2 g_{k_0 i}^4\right). \end{aligned}$$

(ii) Note that

$$\begin{aligned} \widetilde{\mu}_{kl} &= E_{P_2}\left(\widetilde{\zeta}_{(kl)} - \widetilde{\eta}_{(kl)} | \mathcal{F}_k\right) \\ &= E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\left(\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right) + \left(\frac{X_{k,l}}{n/2^k} - 1\right)\tilde{\eta}_{(kl)} | \mathcal{F}_k\right) \\ &= \frac{X_{k,l}}{n/2^k} E_{P_2}\left(\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right) + \left(\frac{X_{k,l}}{n/2^k} - 1\right)\sqrt{n/2^k}\frac{1}{\sqrt{2}}\sqrt{n}\widetilde{g}_{kl}. \end{aligned}$$

A basic inequality $(a+b)^4 \leq 8\left(a^4 + b^4\right)$ gives

$$\begin{aligned} |\widetilde{\mu}_{kl}|^4 \leq\ & 8\left(\frac{X_{k,l}}{n/2^k}\right)^4 \left(E_{P_2}\left(\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right)\right)^4 \\ & + 2\left(\frac{X_{k,l}}{n/2^k} - 1\right)^4 \left(\sqrt{n/2^k}\sqrt{n}\widetilde{g}_{kl}\right)^4. \end{aligned}$$

¿From lemma 2.17, we have

$$\left(E_{P_2} |\widetilde{\mu}_{kl}|^4\right)^{1/2} \leq C\left(E_{P_2}\left(\frac{n}{2^k}\left(1 - 2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right)\right)^2 + C\left(1 + n g_{kl}^2\right) n \widetilde{g}_{kl}^2.$$

Define

$$\tilde{\eta}_{(kl)}^c = \tilde{\eta}_{(kl)}/\sqrt{n/2^k} - \sqrt{\frac{n}{2}}\widetilde{g}_{kl},$$

such that

$$E_{P_2}\tilde{\eta}^c_{(kl)} = 0.$$

Since

$$\frac{n}{2^k}\left(1-2B_{kl}\right)/\sqrt{2} = \sqrt{\frac{n}{2^k}}F_{kl}^{-1}\Phi\left(\tilde{\eta}_{(kl)}/\sqrt{n/2^k}\right)$$

from the equation (2.22), where F_{kl} is the distribution function of $\sqrt{\frac{n}{2^k}}\left(1-2B_{kl}\right)/\sqrt{2}$, then the mean value theorem gives

$$\begin{aligned}
&\frac{n}{2^k}\left(1-2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)} \\
= &\sqrt{\frac{n}{2^k}}\left(F_{kl}^{-1}\Phi\left(\tilde{\eta}_{(kl)}/\sqrt{n/2^k}\right) - \tilde{\eta}_{(kl)}/\sqrt{n/2^k}\right) \\
= &\sqrt{\frac{n}{2^k}}\left(F_{kl}^{-1}\Phi\left(\tilde{\eta}^c_{(kl)}\right) + \left(F_{kl}^{-1}\Phi\left(x\right)\right)'|_{x=\varsigma}\sqrt{\frac{n}{2}}\tilde{g}_{kl} - \tilde{\eta}_{(kl)}/\sqrt{n/2^k}\right),
\end{aligned}$$

where ς is between $\tilde{\eta}^c_{(kl)}$ and $\tilde{\eta}_{(kl)}/\sqrt{n/2^k}$, thus we have

$$\begin{aligned}
E_{P_2}\left(\frac{n}{2^k}\left(1-2B_{kl}\right)/\sqrt{2} - \tilde{\eta}_{(kl)}\right) &= \sqrt{\frac{n}{2^k}}\cdot\sqrt{\frac{n}{2}}\tilde{g}_{kl}E_{P_2}\left(\left(F_{kl}^{-1}\Phi\left(x\right)\right)'|_{x=\varsigma} - 1\right) \\
&= \sqrt{\frac{n}{2^k}}\cdot\sqrt{\frac{n}{2}}\tilde{g}_{kl}E_{P_2}\left(\frac{\varphi\left(\varsigma\right)}{f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)} - 1\right).
\end{aligned}$$

Since $\sqrt{\frac{n}{2}}\tilde{g}_{kl} \ll \left(n/2^k\right)^{1/6}$,i.e., $n^2 2^k \tilde{g}^6_{kl} \ll 1$, the condition $\left|\tilde{\eta}_{(kl)}\right| \le \varepsilon\left(n/2^k\right)^{2/3}$ implies $\left|\tilde{\eta}^c_{(kl)}\right| \le \varepsilon_1\left(n/2^k\right)^{1/6}$ for some $\varepsilon_1 > 0$, then $\left|\varsigma\right| \le \max\left\{\varepsilon, \varepsilon_1\right\}\left(n/2^k\right)^{1/6}$. On the event

$$\widetilde{G}_{kl} = \left\{\left|X_{k,l} - n/2^k\right| \le \left(n/2^k\right)^{2/3},\ \left|\tilde{\eta}_{(kl)}\right| \le \varepsilon\left(n/2^k\right)^{2/3}\right\},$$

we have

$$\frac{\varphi\left(\varsigma\right)}{f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)} = \frac{\varphi\left(\varsigma\right)}{\varphi\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)}\exp\left(O\left(\frac{\varsigma^4}{n/2^k} + \frac{1}{n/2^k}\right)\right),$$

since the formula for f_{kl} is

$$f_{kl}\left(x\right) = \varphi\left(x\right)\exp\left(O\left(\frac{x^4}{n/2^k} + \frac{1}{n/2^k}\right)\right), \text{ for } |x| \le \varepsilon\sqrt{n/2^k}$$

from (2.39). Form the equation (2.23), on the event $\widetilde{G}_{kl}$ we have

$$\begin{aligned} \left|F_{kl}^{-1}\Phi\left(\varsigma\right)^{2}-\varsigma^{2}\right| &= \left|F_{kl}^{-1}\Phi\left(\varsigma\right)-\varsigma\right|\cdot\left|F_{kl}^{-1}\Phi\left(\varsigma\right)-\varsigma+2\varsigma\right| \\ &\leq C_{1}\frac{1}{n/2^{k}}\left(1+\left|\varsigma\right|^{3}\right)\left(2\left|\varsigma\right|+\frac{1}{n/2^{k}}\left(1+\left|\varsigma\right|^{3}\right)\right) \\ &\leq C_{2}\frac{1}{\left(n/2^{k}\right)}\left(1+\left|\varsigma\right|+\left|\varsigma\right|^{4}\right), \end{aligned}$$

which is bounded, then $\varphi\left(\varsigma\right)/f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)$ is bounded on he event $\widetilde{G}_{kl}$. Applying the basic inequality $|x-y|\leq(x+y)\left|\log\left(x/y\right)\right|$ for all positive x and y again, we have

$$\begin{aligned} \left|\frac{\varphi\left(\varsigma\right)}{f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)}-1\right| &\leq \left|\log\varphi\left(\varsigma\right)-\log f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)\right|\left(\frac{\varphi\left(\varsigma\right)}{f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)}+1\right) \\ &\leq C_{3}\left|\log\varphi\left(\varsigma\right)-\log f_{kl}\left(F_{kl}^{-1}\Phi\left(\varsigma\right)\right)\right| \\ &\leq C_{4}\frac{1}{\left(n/2^{k}\right)}\left(1+\left|\varsigma\right|+\left|\varsigma\right|^{4}\right). \end{aligned}$$

Thus we have

$$\begin{aligned} &\left[E_{P_{2}}\left(\frac{n}{2^{k}}\left(1-2B_{kl}\right)/\sqrt{2}-\tilde{\eta}_{(kl)}\right)^{4}\right]^{1/2} \\ \leq\;& C_{1}\left(\sqrt{\frac{n}{2^{k}}}\cdot\sqrt{n}\widetilde{g}_{kl}\right)^{2}\left[\begin{array}{c} E_{P_{2}}\left(\frac{1}{n/2^{k}}\left(1+\left|\varsigma\right|+\left|\varsigma\right|^{4}\right)\right)^{4} \\ +E_{P_{2}}\left(\frac{n}{2^{k}}\left(1-2B_{kl}\right)/\sqrt{2}-\tilde{\eta}_{(kl)}\right)^{4}1_{\widetilde{G}_{kl}^{c}} \end{array}\right]^{1/2} \\ \leq\;& C_{2}\left(\sqrt{n}\widetilde{g}_{kl}\right)^{2}\frac{1}{n/2^{k}}\left(1+\sqrt{n}\widetilde{g}_{kl}+\left(\sqrt{n}\widetilde{g}_{kl}\right)^{8}\right) \\ \leq\;& C_{3}n\widetilde{g}_{kl}^{2}\left(1+n\widetilde{g}_{kl}^{2}\right), \end{aligned}$$

where the last two steps follow from

$$\begin{aligned} &E_{P_{2}}\left(\frac{n}{2^{k}}\left(1-2B_{kl}\right)/\sqrt{2}-\tilde{\eta}_{(kl)}\right)^{4}1_{\widetilde{G}_{kl}^{c}} \\ \leq\;& \left(E\left(\frac{n}{2^{k}}\left(1-2B_{kl}\right)/\sqrt{2}-\tilde{\eta}_{(kl)}\right)^{12}\right)^{1/3} \\ &\cdot\left(E\exp\left(\frac{3}{\sqrt{2}}2^{k_{0}/2}g_{k_{0}i}\eta_{(k_{0}i)}\right)\right)^{1/3}\left(E1_{\widetilde{G}_{kl}^{c}}\right)^{1/3} \\ \leq\;& C_{1}\exp\left(-C_{2}\left(n/2^{k}\right)^{1/3}\right) \end{aligned}$$

by similar arguments in the lemma 2.3, and

$$\sqrt{n}\widetilde{g}_{kl} \ll \left(n/2^k\right)^{1/6}.$$

So we have

$$\left(E_{P_2}\widetilde{\mu}^4_{(kl)}\right)^{1/2} \leq C\left(1+ng^2_{kl}+n\widetilde{g}^2_{kl}\right)n\widetilde{g}^2_{kl}.$$

(iii) Cauchy-Shwartz inequality yields

$$E\left(V_i+1\right)^4 U_i^4 \Lambda_{2,i,n} \leq \left(E\left(V_i+1\right)^8 \Lambda_{2,i,n}\right)^{1/2}\left(EU_i^8\Lambda_{2,i,n}\right)^{1/2}$$

¿From lemma (2.2) $\left(EU_i^8\Lambda_{2,i,n}\right)^{1/2}$ is uniformly bounded, then it is enough to show $EV_i^m\Lambda_{2,i,n}$ is uniformly bounded for a positive number $m>0$. Hölder inequality gives

$$\begin{aligned} E_{P_2}V_i^m &\leq \left(E_{P_2}\exp\left(K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)}\right)\right)^{1/K} \\ &\cdot\prod_{k=k_0}^{k_1-1}\left(\prod_{l\in I_k^i}E_{P_2}\left[\exp\left(K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)}\right)\right]\right)^{1/K}. \end{aligned}$$

The mean values theorem yields

$$\begin{aligned} \exp\left(K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)}\right) &= 1+K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)}\exp\left(K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\varsigma_{(k_0 i)}\right) \\ \exp\left(K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)}\right) &= 1+K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)}\exp\left(K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\widetilde{\varsigma}_{(kl)}\right) \end{aligned}$$

where $\left|\varsigma_{(k_0 i)}\right| \leq \left|\mu_{(k_0 i)}\right|$ and $\left|\widetilde{\varsigma}_{(kl)}\right| \leq \left|\widetilde{\mu}_{(kl)}\right|$, then applying Cauchy-Shwartz inequality, and combining with remark 2.4 and the first two inequalities of lemma 2.4, we have

$$\begin{aligned} & E_{P_2}\exp\left(K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)}\right) \\ =\ & 1+E_{P_2}K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)}\exp\left(K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\varsigma_{(k_0 i)}\right) \\ \leq\ & 1+\left|K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\right|\left(E_{P_2}\mu^2_{(k_0 i)}\right)^{1/2}\left[E_{P_2}\exp\left(2\left|K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\mu_{(k_0 i)}\right|\right)\right]^{1/2} \\ \leq\ & 1+C\left|K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\right|\left(1+n^2g^4_{k_0 i}\right)^{1/2} \end{aligned}$$

and

$$
\begin{aligned}
& E_{P_2} \exp\left(K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)}\right) \\
= \ & 1 + E_{P_2} K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)} \exp\left(K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\varsigma}_{(kl)}\right) \\
\leq \ & 1 + \left|K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right| \left(E_{P_2}\widetilde{\mu}^4_{(kl)}\right)^{1/4} \left[E_{P_2}\exp\left(2\left|K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\widetilde{\mu}_{(kl)}\right|\right)\right]^{1/2} \\
\leq \ & 1 + C\left|K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right| \left(n\widetilde{g}^2_{kl}\left(1 + ng^2_{kl} + n\widetilde{g}^2_{kl}\right)\right)^{1/2}.
\end{aligned}
$$

Thus applying Cauchy-Schwartz inequality twice yields

$$
\begin{aligned}
& \left(\log E_{P_2} V_i^m\right)^2 \\
\leq \ & C\left[\frac{1}{K}\left(\begin{array}{c} \left|K\frac{m}{\sqrt{2}}2^{k_0/2}g_{k_0 i}\right| \left(1 + n^2 g^4_{k_0 i}\right)^{1/2} \\ +\sum_{k=k_0}^{k_1-1}\sum_{l\in I_k^i}\left|K\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right|\left(n\widetilde{g}^2_{kl}\left(1+ng^2_{kl}+n\widetilde{g}^2_{kl}\right)\right)^{1/2}\end{array}\right)\right]^2 \\
\leq \ & K\left(\begin{array}{c} 2^{k_0}g^2_{k_0 i}\left(1 + n^2 g^4_{k_0 i}\right) \\ +\sum_{k=k_0}^{k_1-1}\left(\sum_{l\in I_k^i}\left|\frac{m}{\sqrt{2}}2^{k/2}\ \widetilde{g}_{kl}\right|\left(n\widetilde{g}^2_{kl}\left(1+ng^2_{kl}+n\widetilde{g}^2_{kl}\right)\right)^{1/2}\right)^2\end{array}\right) \\
\leq \ & C\log n\left(2^{k_0}g^2_{k_0 i}\left(1+n^2g^4_{k_0 i}\right) + \sum_{k=k_0}^{k_1-1}\sum_{l\in I_k^i}\frac{2^{2k}}{2^{k_0}}\ \widetilde{g}^2_{kl}n\widetilde{g}^2_{kl}\left(1+ng^2_{kl}+n\widetilde{g}^2_{kl}\right)\right)
\end{aligned}
$$

which is uniformly bounded from the equation (2.32), i.e., $EV_i^m\Lambda_{2,i,n}$ is uniformly bounded. So this inequality is established.

Combining (2.15), (2.30), (2.31) and lemma 2.13 we have

Theorem 2.2 *Under the assumption For $f \in \Sigma_n(f_0)$, we have*

$$
\begin{aligned}
& H^2\left(\Lambda_{1,n}, \Lambda_{2,n}\right) \\
\leq \ & C(\log n)\left[n^{3/2}\int|f-f_0|^6 + n^{1/2}\int|f-f_0|^2 + \sum_{k=k_0}^{k_1-1}\sum_l\left(2^k\ \widetilde{g}^2_{kl} + 2^{5k}\ \widetilde{g}^6_{kl}\right)\right]
\end{aligned}
$$

goes to zero uniformly.

2.5 The Preliminary Estimator

We have shown the local asymptotic equivalence between $E^1_{1,n}$ and $E^1_{2,n}$ by theorem 2.2. In this section we construct the preliminary estimators in two experiments $E^1_{1,n}$ and $E^1_{2,n}$, which imply " nonparametric " neighborhoods (2.8) in section 2.2 are attainable such that the local approximations can be glued together in a global approximation.

Note that $E^1_{1,n}$ and $E^1_{2,n}$ are the following regression type model

$$Y_j = J_{j,n}(g) + \zeta_j, \ j = 1, 2, ..., n \tag{2.33}$$

where ζ_j are known noise. Let's assume $E\ \zeta_j = 0$, otherwise we translate the model with known location.

The following lemma gives the construction of preliminary estimators for regression model (2.33).

Lemma 2.5 *For the following regression type model*

$$Y_j = J_{j,n}(g) + \zeta_j, j = 1, 2, ..., n,$$

where $J_{j,n}(g) = n\int_{(j-1)/n}^{j/n} g(x)\,dx$, *and* ξ_j *are i.i.d. 0 mean with pth moments, and* g *is in a compact set of* $B^{\alpha}_{p,p}$, *for* $\alpha - 1/p + 1/2 > 0$, $p \geq 1$, *then there is an estimator* $\widehat{g}_n$ *such that*

$$\sup_g E_g \left\| g - \widehat{g}_n \right\|_p = o\left(n^{-\alpha/(2\alpha+1)}\right).$$

Proof: Consider the averaging operator

$$J_{j,n}(g) = n\int_{(j-1)/n}^{j/n} g(x)dx.$$

We observe the regression type model

$$Y_j = J_{j,n}(g) + \zeta_j,\ j = 1, \ldots, n$$

where ξ_j are i.i.d. 0 mean with p moments. For some $p \geq 1$, consider the expression

$$\sum_{j=1}^{n} \left(n \int_{(j-1)/n}^{j/n} g(x)dx \right)^p = \sum_{j=1}^{n} J_{j,n}^p(g) = n \|g\|_{(n),p}^p \ ,$$

where

$$\|g\|_{(n),p} = \left(n^{-1} \sum_{j=1}^{n} J_{j,n}^p(g) \right)^{1/p}$$

is a seminorm. Note that due to Jensen's inequality

$$(J_{j,n}(g))^p \leq n \int_{(j-1)/n}^{j/n} |g|^p (x)dx$$

so that

$$\begin{aligned} \|g\|_{(n),p}^p &= n^{-1} \sum_{j=1}^{n} J_{j,n}^p(g) \leq \sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} |g|^p (x)d \\ &= \int_0^1 |g|^p (x)dx = \|g\|_p^p . \end{aligned}$$

The question is now whether we can find an estimator $\widehat{g}_n$ such that

$$E_g \|\widehat{g}_n - g\|_p = o\left(n^{-\alpha/(2\alpha+1)}\right)$$

uniformly over all g,which implies

$$E_g \|\widehat{g}_n - g\|_{(n),p} = o\left(n^{-\alpha/(2\alpha+1)}\right)$$

uniformly over all g. We shall use a piecewise constant (histogram type) estimator $\widehat{g}_{(m)}$ with smoothing parameter m where m^{-1} is the bin width.

Assume that $s = n/m$ is integer. Let $\widehat{g}_{(m)}$ be

$$\widehat{g}_{(m)} = \sum_{k=1}^{m} \hat{J}_{k,m} \mathbf{1}_{[(k-1)/m, k/m)}$$

where

$$\hat{J}_{k,m} = s^{-1} \sum_{j=(k-1)s+1}^{ks} Y_j.$$

The estimator can be decomposed

$$\widehat{g}_{(m)} = E\widehat{g}_{(m)} + \widehat{g}_{(m)} - E\widehat{g}_{(m)}$$

with a corresponding risk decomposition

$$E_g \left\|\widehat{g}_{(m)} - g\right\|_p \leq \left\|E\widehat{g}_{(m)} - g\right\|_p + E_g \left\|\widehat{g}_{(m)} - E\widehat{g}_{(m)}\right\|_p .$$

Now

$$E_g \hat{J}_{k,m} = s^{-1} \sum_{j=(k-1)s}^{ks} J_{j,n}(g) = J_{k,m}(g)$$

so that $E\widehat{g}_{(m)}$ is a linear approximation operator

$$E\widehat{g}_{(m)} = \Pi_m g = \sum_{j=1}^{m} J_{j,m}(g) \mathbf{1}_{[(j-1)/m, j/m)},$$

where Π_m is a projection operator. For the approximation operator Π_m we will show (setting $m = 2^r$ for some $r \to \infty$)

$$\|\Pi_m g - g\|_p = o\left(m^{-\alpha}\right)$$

uniformly over all g. For $p > 1$ we have

$$\|\Pi_m g - g\|_p^p = \int_0^1 \left(\sum_{i \geq r} \sum_{i=1}^{2^i} \tilde{g}_{ij} \psi_{ij} \right)^p \leq \int_0^1 R_1 R_2$$

where

$$R_1 = \left(\sum_{i \geq r} \left(2^{-\alpha i}\right)^{p/(p-1)} \right)^{p-1}, \; R_2 = \sum_{i \geq r} \left| \sum_{j=1}^{2^i} 2^{\alpha i} \tilde{g}_{ij} \psi_{ij} \right|^p$$

from Hölder inequality $\left|\sum_{i \geq r} a_i b_i\right| \leq \left(\sum_{i \geq r} |a_i|^{p/(p-1)}\right)^{(p-1)/p} \left(\sum_{i \geq r} |b_i|^p\right)^{1/p}$. We see

$$R_1 \leq C 2^{-\alpha p r},$$

and

$$\int_0^1 R_2 \leq \sum_{i \geq r} 2^{p\alpha i} \int_0^1 \left| \sum_{j=1}^{2^i} \widetilde{g}_{ij} \psi_{ij} \right|^p = \sum_{i \geq r} \sum_{j=1}^{2^i} 2^{p(\alpha + 1/2 - 1/p)i} \left|\widetilde{g}_{ij}\right|^p = o\,(1)$$

uniformly from the compactness of the parameter space in $B^{\alpha}_{p,p}$, thus

$$\|\Pi_m g - g\|_p^p = o\left(m^{-\alpha}\right)$$

uniformly over all g.

It remains to show

$$E_f \left\|\widehat{g}_{(m)} - E\widehat{g}_{(m)}\right\|_p \le C\left(\frac{m}{n}\right)^{1/2}.$$

If this is true, then the usual optimal choice of m gives

$$E_f \left\|\widehat{g}_n - g\right\|_p = o\left(n^{-\alpha/(2\alpha+1)}\right).$$

Note that for $s = n/m$ (assumed integer) and $t \in [0, m^{-1}]$ we have

$$\widehat{g}_{(m)}(t) - \Pi_m g(t) = s^{-1}\sum_{j=1}^{s}\zeta_j$$

and for t in another interval $[(k-1)m^{-1}, km^{-1}]$ we have an analogous average of s independent noise terms ζ_j. Denote these averages

$$\eta_k = s^{-1}\sum_{j=(k-1)s}^{ks}\zeta_j \text{ for } k = 1, \ldots, m.$$

Let us now treat the case of the general L_p norm. Thus we obtain

$$\begin{aligned} E_f \left\|\widehat{g}_{(m)} - E\widehat{g}_{(m)}\right\|_p &= E\left(m^{-1}\sum_{k=1}^{m}|\eta_k|^p\right)^{1/p} \\ &\le \left(m^{-1}\sum_{k=1}^{m}E\,|\eta_k|^p\right)^{1/p} \end{aligned}$$

where the last inequality holds by concavity of $x \to x^{1/p}$ for $p \ge 1$. Now by an inequality of Dharmadikari and Jogdeo (also cited as Whittle's inequality), (cf Petrov, p. 60, Sec. III.5, probl. 16)

$$E\left|\sum_{j=1}^{s}\zeta_j\right|^p \le C_p s^{p/2-1}\sum_{j=1}^{s}E\left|\zeta_j\right|^p = C_p s^{p/2}E\left|\zeta_1\right|^p$$

and assuming $E\left|\zeta_1\right|^p < \infty$ we obtain

$$E\left|\eta_k\right|^p \leq Cs^{-p}s^{p/2} = Cs^{-p/2} = C\left(\frac{m}{n}\right)^{p/2}.$$

This proves that

$$\left(m^{-1}\sum_{k=1}^{m} E\left|\eta_k\right|^p\right)^{1/p} \leq C\left(\left(\frac{m}{n}\right)^{p/2}\right)^{1/p} = C\left(\frac{m}{n}\right)^{1/2}.$$

Remark 2.5 *Lemma 2.8 tells us*

$$\Sigma' = \{\log f : f \in \Sigma\} \subset B^{\alpha}_{2,2}(M') \cap B^{\alpha}_{6,6}(M'), \textit{ for some } M' > 0.$$

then there is an estimator $\hat{f}_n$ *such that*

$$\sup_f E_f \left\|\log \widehat{f}_n - \log f\right\|_6 = o\left((\log n)^{-2} n^{-1/4}\right)$$

in experiment $E^1_{1,n}$ *(or* $E^1_{2,n}$*). From the compactness of* Σ' *in* L_6 *metric, we can assume the estimator belong to* Σ *similar to Lemma 3.1 in Nussbaum (1996)., thus we have*

$$\sup_f E_f \left\|\widehat{f}_n - f\right\|_6 \leq C\sup_f E_f \left\|\log \widehat{f}_n - \log f\right\|_6 = o\left((\log n)^{-2} n^{-1/4}\right).$$

2.6 Quantile Coupling for Beta Distributions

The quantile coupling method was introduced in Komlós, Major, and Tusnády(1975), which is considered one of the most important probability papers of the last forty years, and now often referred to as the KMT. Their coupling greatly simplifies the derivation of many classical results (see Shorack & Wellner (1986)), and is a powerful way to obtain strong approximation results. Here we establish a quantile coupling bound between a Beta random variable and a normal random variable, which improves the classical bound in KMT paper with a rate. This result helps us to have close matchings between $\widetilde{\zeta}_{kl}$ and $\widetilde{\eta}_{kl}$ in (2.17), and stated in (2.23), (2.24), and (2.25).

Lemma 2.6 *Let Z be a standard normal random variable. For every n, there is a mapping $T_n : R \mapsto R$ such that the random variable $B_n = T_n(Z)$ has the Beta $(n/2, n/2)$ law and*

$$\left| n\left(1/2 - B_n\right) - \frac{n^{1/2}}{2} Z \right| \leq \frac{C}{\sqrt{n}} + \frac{C}{n^2} \left| nB_n - n/2 \right|^3 \tag{2.34}$$

for $|nB_n - n/2| \leq \varepsilon n$, where C, $\varepsilon > 0$ do not depend on n.

Remark 2.6 *Since Z is symmetric, there is also a mapping $T'_n : R \mapsto R$ such that the random variable $B_n = T'_n(Z)$ has the Beta $(n/2, n/2)$ law and*

$$\left| n\left(B_n - 1/2\right) - \frac{n^{1/2}}{2} Z \right| \leq \frac{C}{\sqrt{n}} + \frac{C}{n^2} \left| nB_n - n/2 \right|^3$$

for $|nB_n - n/2| \leq \varepsilon n$, where C, $\varepsilon > 0$ do not depend on n.

Proof: This proof has two steps. In step I we approximate the distribution function of B_n by the distribution function of normal random variable. Based on the approximation of distribution functions we derive the sharp inequality for quantile coupling in step II.

Step I: Let B_n have the Beta $(n/2, n/2)$ law, $n = 2k + 2$. We have

$$EB_n = 1/2,\ \operatorname{Var}\left(B_n\right) = \frac{1}{4(2k+3)}.$$

The density function of $B_n^* = 1/2 - B_n$ is

$$g\left(x\right) = \frac{(2k+1)!}{\left(k!\right)^2} \left(\frac{1}{2} + x\right)^k \left(\frac{1}{2} - x\right)^k = \frac{(2k+1)!}{\left(k!\right)^2} \exp\left(k \log\left(\frac{1}{4} - x^2\right)\right).$$

We apply Stirling formula, $j! = \sqrt{2\pi} j^{j+1/2} \exp(-j+\epsilon_j)$ with $\epsilon_j = O(1/j)$, to $(2k+1)!$ and $k!$, then give

$$g(x) = \frac{2}{\sqrt{2\pi}} \sqrt{2k+1} \left(\frac{2k+1}{2k}\right)^{2k+1} \exp(-1) \exp\left(k \log\left(1-4x^2\right)\right) \exp\left(O\left(\frac{1}{k}\right)\right). \tag{2.35}$$

Since $|\log(1+x) - x| \leq 2x^2$ when $|x| < 1/2$, we have

$$\left(\frac{2k+1}{2k}\right)^{2k+1} = \exp\left(-(2k+1)\log\left(1-\frac{1}{2k+1}\right)\right) = \exp\left(1+O\left(\frac{1}{k}\right)\right). \tag{2.36}$$

It is easy to see

$$\left|\frac{\sqrt{2k+1}}{\sqrt{2k}} - 1\right| = O\left(\frac{1}{k}\right). \tag{2.37}$$

Combining (2.35), (2.36) and (2.37), we have

$$g(x) = \frac{\sqrt{8k}}{\sqrt{2\pi}} \exp\left(k \log\left(1-4x^2\right)\right) \exp\left(O\left(\frac{1}{k}\right)\right). \tag{2.38}$$

Furthermore Taylor expansion gives

$$\log\left(1-4x^2\right) = -4x^2 + x^4 \lambda(x), \text{ for } |x| \leq 2\varepsilon \text{ and } 0 < 2\varepsilon < 1/2,$$

where $\lambda(x)$ is an analytic function, which is uniformly bounded. Thus we have

$$\begin{aligned} g(x) &= \frac{\sqrt{8k}}{\sqrt{2\pi}} \exp\left(-4kx^2 + kx^4\lambda(x)\right) \exp\left(O\left(k^{-1}\right)\right) \\ &= \sqrt{8k}\varphi\left(\sqrt{8k}x\right) \exp\left(kx^4\lambda(x) + O\left(k^{-1}\right)\right), \text{ for } |x| \leq 2\varepsilon, \end{aligned} \tag{2.39}$$

which is a delicate approximation of the density function of B_n by the density function of normal random variable. Now we want to use this approximation of density functions to give the desired approximation of distribution functions, and show

$$\begin{aligned} G(x) &= \int_{-1/2}^{x} g(t) \\ &\leq \Phi\left(\sqrt{8k}x\right) \exp\left(Ckx^4 + Ck^{-1}\right) \end{aligned} \tag{2.40}$$

and

$$G(x) \geq \Phi\left(\sqrt{8k}x\right) \exp\left(-Ckx^4 - Ck^{-1}\right) \tag{2.41}$$

for any $-\varepsilon \leq x \leq 0$ and some $C > 0$. The formulation for $0 \leq x \leq \varepsilon$ is similar, and will be given later in (2.43). Now we give the proof for the inequality (2.40). The equation (2.41) follows similarly. In (2.38) it is easy to see

$$g(a)/g(b) = \frac{\sqrt{8k}}{\sqrt{2\pi}} \exp\left(k \log\left[\left(1-4a^2\right)/\left(1-4b^2\right)\right]\right) = o\left(k^{-1}\right), \text{ for } |a|-|b| > \varepsilon/2$$

which implies

$$\int_{-1/2}^{-2\varepsilon} g(t) = O\left(k^{-1}\right) \int_{-3\varepsilon/2}^{x} g(t) = O\left(k^{-1}\right) \int_{-2\varepsilon}^{x} g(t),$$

i.e.,

$$G(x) = \left(1 + O\left(k^{-1}\right)\right) \int_{-2\varepsilon}^{x} g(t).$$

Note that

$$\begin{aligned} & \left(\Phi\left(\sqrt{8k}x\right) \exp\left(Ckx^4 + Ck^{-1}\right)\right)' \\ = \; & \sqrt{8k}\varphi\left(\sqrt{8k}x\right) \exp\left(Ckx^4 + Ck^{-1}\right) \\ & + \Phi\left(\sqrt{8k}x\right)\left(4Ckx^3\right) \exp\left(Ckx^4 + Ck^{-1}\right). \end{aligned} \tag{2.42}$$

The inequality

$$\Phi\left(\sqrt{8k}x\right)\left(-\sqrt{8k}x\right) < \varphi\left(\sqrt{8k}x\right)$$

yields

$$\begin{aligned} & \Phi\left(\sqrt{8k}x\right)\left(4Ckx^3 + 2Cx\right) \exp\left(Ckx^4 + Ck^{-1}\right) \\ \geq \; & -\sqrt{8k}\varphi\left(\sqrt{8k}x\right)\left(\frac{C}{2}x^2\right) \exp\left(Ckx^4 + Ck^{-1}\right), \end{aligned}$$

then combining with (2.42) we have

$$\begin{aligned}
& \left(\Phi\left(\sqrt{8k}x\right)\exp\left(Ckx^4+C\left(x^2+k^{-1}\right)\right)\right)' \\
\geq\ & \sqrt{8k}\varphi\left(\sqrt{8k}x\right)\left(1-\frac{C}{2}x^2\right)\exp\left(Ckx^4+Ck^{-1}\right) \\
\geq\ & \sqrt{8k}\varphi\left(\sqrt{8k}x\right)\exp\left(-Cx^2\right)\exp\left(Ckx^4+Ck^{-1}\right) \\
=\ & \sqrt{8k}\varphi\left(\sqrt{8k}x\right)\exp\left(\frac{C}{2}kx^4+\frac{C}{2}k^{-1}+\frac{C}{2}k\left(x^2-1/k\right)^2\right) \\
\geq\ & \sqrt{8k}\varphi\left(\sqrt{8k}x\right)\exp\left(\frac{C}{2}kx^4+\frac{C}{2}k^{-1}\right),
\end{aligned}$$

where in the second inequality we apply $(1-Cx^2/2)\geq\exp(-Cx^2)$ when $|Cx^2|\leq C(2\varepsilon)^2<1$. Since $\lambda(x)$ is uniformly bounded on $[-2\varepsilon,0]$, the inequality above implies

$$\left(\Phi\left(\sqrt{8k}x\right)\exp\left(Ckx^4+Ck^{-1}\right)\right)'\geq\sqrt{8k}\varphi\left(\sqrt{8k}x\right)\exp\left(kx^4\lambda(x)+O\left(k^{-1}\right)\right)$$

for C sufficiently large. Thus we have

$$\begin{aligned}
G(x) &= \left(1+O\left(k^{-1}\right)\right)\int_{-2\varepsilon}^{x}g(t) \\
&\leq \left(1+O\left(k^{-1}\right)\right)\int_{-2\varepsilon}^{x}\left(\Phi\left(\sqrt{8k}t\right)\exp\left(Ckt^4+Ck^{-1}\right)\right)' \\
&= \exp\left(O\left(k^{-1}\right)\right)\left[\begin{array}{c}\Phi\left(\sqrt{8k}x\right)\exp\left(C\left(kx^4+k^{-1}\right)\right) \\ -\Phi\left(\sqrt{8k}\cdot(2\varepsilon)\right)\exp\left(C\left(k(2\varepsilon)^4+k^{-1}\right)\right)\end{array}\right] \\
&\leq \Phi\left(\sqrt{8k}x\right)\exp\left(C\left(kx^4+k^{-1}\right)\right),
\end{aligned}$$

which is the equation (2.40). Similarly we may establish (2.41). Thus we have

$$G(x)=\Phi\left(\sqrt{8k}x\right)\exp\left(O\left(kx^4+k^{-1}\right)\right),\text{ for } -\varepsilon\leq x\leq 0.$$

Similarly it can be shown

$$1-G(x)=\left(1-\Phi\left(\sqrt{8k}x\right)\right)\exp\left(O\left(kx^4+k^{-1}\right)\right),\text{ for } 0\leq x\leq\varepsilon. \tag{2.43}$$

Step II: Define

$$B_n^* = G^{-1}\Phi\left(Z\right), \text{ i.e., } Z = \Phi^{-1}G\left(B_n^*\right). \tag{2.44}$$

which is equivalent to define

$$\sqrt{4n}B_n^* = \widetilde{G}^{-1}\Phi\left(Z\right), \text{ i.e., } Z = \Phi^{-1}\widetilde{G}\left(\sqrt{4n}B_n^*\right) \tag{2.45}$$

where $\widetilde{G}$ is the distribution function of $\sqrt{4n}B_n^*$. We want to show

$$-C\frac{1}{n}\left(1+(\sqrt{n}\left|B_n^*\right|)^3\right) \leq \sqrt{4n}B_n^* - Z \leq C\frac{1}{n}\left(1+(\sqrt{n}\left|B_n^*\right|)^3\right), \text{ for } \left|B_n^*\right| \leq \varepsilon.$$

Since

$$\left|\sqrt{8k}B_n^* - \sqrt{4n}B_n^*\right| \leq \frac{1}{k}\left|B_n^*\right| \leq \frac{1}{k}\left(1+\left|B_n^*\right|^3\right)$$

and $n \asymp k$, it is equivalent to show that there is a constant C such that

$$-C\frac{1}{k}\left(1+(\sqrt{k}\left|B_n^*\right|)^3\right) \leq \sqrt{8k}B_n^* - Z \leq C\frac{1}{k}\left(1+(\sqrt{k}\left|B_n^*\right|)^3\right), \text{ for } \left|B_n^*\right| \leq \varepsilon,$$

i.e.,

$$\begin{aligned} & \Phi\left(\sqrt{8k}\left|B_n^*\right| - C\frac{1}{k}\left(1+(\sqrt{k}\left|B_n^*\right|)^3\right)\right) \\ \leq \;& G\left(B_n^*\right) \leq \Phi\left(\sqrt{8k}\left|B_n^*\right| + C\frac{1}{k}\left(1+(\sqrt{k}\left|B_n^*\right|)^3\right)\right), \text{ for } \left|B_n^*\right| \leq \varepsilon. \end{aligned}$$

Let's consider the case $0 \leq B_n^* \leq \varepsilon$ only (The derivation is similar for $-\varepsilon \leq B_n^* \leq 0$). It is equivalent to show

$$\begin{aligned} & 1-\Phi\left(\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right) \\ \geq \;& 1-G\left(x\right) \geq 1-\Phi\left(\sqrt{8k}x + C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right) \end{aligned} \tag{2.46}$$

when $0 \le B_n^* \le \varepsilon$, i.e.,

$$\log\left(\frac{1-\Phi\left(\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right)}{1-\Phi\left(\sqrt{8k}x\right)}\right)$$

$$\geq \log\frac{1-G(x)}{1-\Phi\left(\sqrt{8k}x\right)} \tag{2.47}$$

$$\geq \log\left(\frac{1-\Phi\left(\sqrt{8k}x + C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right)}{1-\Phi\left(\sqrt{8k}x\right)}\right) \tag{2.48}$$

when $0 \le B_n^* \le \varepsilon$, where

$$\log\frac{1-G(x)}{1-\Phi\left(\sqrt{8k}x\right)} = O\left(kx^4 + k^{-1}\right).$$

Now we show (2.47). It is easy to see that the first part of the equation (2.46) is satisfied when $\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right) \le 0$, because the left hand side is more than 1/2, while the right hand side is less than 1/2. So (2.47) is satisfied when $\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right) \le 0$. Then we need only to consider the case $\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right) \ge 0$. The intermediate value theorem tells us there is a number ξ between $\sqrt{8k}x$ and $\sqrt{8k}x - \frac{C}{2}\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)$ such that

$$\log\left(\frac{1-\Phi\left(\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right)}{1-\Phi\left(\sqrt{8k}x\right)}\right)$$

$$\geq \log\left(\frac{1-\Phi\left(\sqrt{8k}x - \frac{C}{2}\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right)}{1-\Phi\left(\sqrt{8k}x\right)}\right)$$

$$= \frac{C}{2}\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\frac{\varphi(\xi)}{1-\Phi(\xi)}.$$

¿From the lemma (2.7), we have

$$\begin{aligned}
& \log\left(\frac{1-\Phi\left(\sqrt{8k}x - C\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\right)}{1-\Phi\left(\sqrt{8k}x\right)}\right) \\
\geq\ & \frac{C}{2}\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\cdot\frac{1}{2}\left(\sqrt{8k}x - \frac{C}{2}\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)+\frac{\sqrt{2\pi}}{2}\right) \\
\geq\ & \frac{C}{2}\frac{1}{k}\left(1+(\sqrt{k}x)^3\right)\cdot\frac{1}{2}\left(\frac{1}{2}\sqrt{8k}x+\frac{\sqrt{2\pi}}{2}\right) \\
\geq\ & \frac{C}{\sqrt{8}}kx^4+\frac{C\sqrt{2\pi}}{4k},
\end{aligned}$$

which is bigger than the right side of the equation (2.47) for C sufficiently large. Thus we establish (2.47). The equation (2.48) can be established similarly.

Lemma 2.7 *For* $x > 0$

$$\bar{\Phi}(x) < \min\left\{\frac{1}{x}, \frac{\sqrt{2\pi}}{2}\right\}\varphi(x)$$

i.e.

$$\frac{\varphi(x)}{\bar{\Phi}(x)} > \min\left\{x, \frac{2}{\sqrt{2\pi}}\right\} \geq \frac{1}{2}\left(x+\frac{2}{\sqrt{2\pi}}\right).$$

Corollary 2.1 *We have*

$$\left|nB_n - n/2 - \left(n^{1/2}/2\right)Z\right| \leq \frac{C}{\sqrt{n}}\left(1+|Z|^3\right), \text{ when } |Z| \leq \varepsilon\sqrt{n}$$

for some $C, \varepsilon > 0$.

Proof: Obviously the inequality (2.34) still holds, when $|nB_n - n/2| \leq \varepsilon_1 n$ for $0 < \varepsilon_1 \leq \varepsilon$. Let's choose ε_1 small enough such that $C\varepsilon_1^2 < 1/2$. When $|nB_n - n/2| \leq \varepsilon_1 n$, we have

$$\left|nB_n - n/2 - \frac{n^{1/2}}{2}Z\right| \leq \frac{C}{\sqrt{n}} + \frac{1}{2}\left|nB_n - n/2\right|,$$

from (2.34), which implies

$$|nB_n - n/2| - \left|\frac{n^{1/2}}{2}Z\right| \le \frac{C}{\sqrt{n}} + \frac{1}{2}\,|nB_n - n/2|$$

by the triangle inequality, i.e.,

$$|nB_n - n/2| \le \frac{2C}{\sqrt{n}} + \left|n^{1/2}Z\right|, \tag{2.49}$$

so we have

$$\left|nB_n - n/2 - \frac{n^{1/2}}{2}Z\right| \le \frac{C}{\sqrt{n}} + \frac{C}{n^2}\left(\frac{C}{\sqrt{n}} + \left|\frac{n^{1/2}}{2}Z\right|\right)^3 \le \frac{C_1}{\sqrt{n}}\left(1 + |Z|^3\right)$$

for some $C_1 > 0$.

When $nB_n - n/2 = \varepsilon_1 n > 0$, we know $Z \ge 0$ from the definition of quantile coupling in (2.44), and from the equation (2.49) we have

$$n^{1/2}Z \ge \varepsilon_1 n - \frac{2C}{\sqrt{n}}.$$

In the definition of quantile coupling, we see that $nB_n - n/2$ is an increasing function of Z. So we have $nB_n - n/2 \le \varepsilon_1 n$, when $n^{1/2}Z \le \varepsilon_1 n - \frac{2C}{\sqrt{n}}$. Similarly we may show $nB_n - n/2 \ge -\varepsilon_1 n$, when $n^{1/2}Z \ge -\varepsilon_1 n - \frac{2C}{\sqrt{n}}$. Thus we have

$$|nB_n - n/2| \le \varepsilon_1 n, \text{ when } \left|n^{1/2}Z\right| \le \varepsilon_1 n - \frac{2C}{\sqrt{n}}. \tag{2.50}$$

Let $\varepsilon_2 = \varepsilon_1/2$. We have $\varepsilon_2\sqrt{n} < \varepsilon_1\sqrt{n} - \frac{2C}{n}$ for $n > \left(\frac{2C}{\varepsilon_2}\right)^{2/3}$, then we know

$$\left\{|Z| \le \varepsilon_2\sqrt{n}\right\} \subset \left\{\ |Z| \le \varepsilon_1\sqrt{n} - \frac{2C}{n}\right\} \subset \left\{|nB_n - n/2| \le \varepsilon_1 n\right\}$$

from (2.50), so we have

$$\left|nB_n - n/2 - \frac{n^{1/2}}{2}Z\right| \le \frac{C_1}{\sqrt{n}}\left(1 + |Z|^3\right), \text{ when } |Z| \le \varepsilon_2\sqrt{n} \text{ and } n > \left(\frac{2C}{\varepsilon_2}\right)^{2/3}.$$

Obviously we have

$$\left|nB_n - n/2 - \frac{n^{1/2}}{2}Z\right| \le \frac{C_2}{\sqrt{n}}$$

for some$C_2 > 0$, when $|Z| \leq \varepsilon_2\sqrt{n}$ and $n < \left(\frac{2C}{\varepsilon_2}\right)^{2/3}$. Let $C_3 = \max\{C_1, C_2\}$, then we have

$$\left|nB_n - n/2 - \frac{n^{1/2}}{2}Z\right| \leq \frac{C_3}{\sqrt{n}}\left(1 + |Z|^3\right), \text{ when } |Z| \leq \varepsilon_2\sqrt{n}.$$

Remark 2.7 *A closed related result is Carter and Pollard (2004) which improved Tusnády's inequality, modulo constants. For the coupling between an X distributed $Bin(n, 1/2)$ and a $Y = n/2 + \sqrt{n}Z/2$ distributed $N(n/2, n/4)$, they show*

$$|X - Y| \leq C + C\frac{|Z|^3}{\sqrt{n}}, \text{ when } |Z| \leq \varepsilon\sqrt{n}$$

for some $C, \varepsilon > 0$. In our theorem, the upper bound is $C\left(1 + |Z|^3\right)/\sqrt{n}$, which is in a level of $1/\sqrt{n}$ if we see Z as a constant level.

2.7 Appendix

In the course of the reasoning we made use of the following simple auxiliary results.

The proof of the following lemma is similar to that of lemmas 1 and 2 in the appendix of Brown, Carter, Low and Zhang (2004).

Lemma 2.8 *Let $f \in \Sigma$. Then (i)*

$$\left|2^{k/2}(\log f)_{kl} - \log\left(2^{k/2}f_{kl}\right)\right| \leq C2^k \int_{I_{kl}} \left(f - 2^{k/2}f_{kl}\right)^2.$$

(ii)

$$\{\log f : f \in \Sigma\} \subset B^{\alpha}_{2,2}(M') \cap B^{\alpha}_{6,6}(M'), \text{ for some } M' > 0.$$

Proof: (i) Taylor expansion gives

$$\log\frac{f}{2^{k/2}f_{kl}} = \frac{f}{2^{k/2}f_{kl}} - 1 - \frac{1}{2a^2}\left(\frac{f}{2^{k/2}f_{kl}} - 1\right)^2,$$

where a is between 1 and $f/\left(2^{k/2}f_{kl}\right)$. The assumption $f \in \Sigma \subset \mathcal{F}_\epsilon$ implies a is bounded away from 0. Since $\int_{I_{kl}} f/\left(2^{k/2}f_{kl}\right) = 1$ from the definition of f_{kl} in (2.3), then

$$\begin{aligned} & \left|2^{k/2}\left(\log f\right)_{kl} - \log\left(2^{k/2}f_{kl}\right)\right| \\ = & \left|2^k \int_{I_{kl}} \log \frac{f}{2^{k/2}f_{kl}}\right| \\ = & 2^k \int_{I_{kl}} \frac{1}{2a^2}\left(1 - \frac{f}{2^{k/2}f_{kl}}\right)^2 \le C2^k \int_{I_{kl}} \left(f - 2^{k/2}f_{kl}\right)^2. \end{aligned}$$

(ii) Since $\widetilde{\log f}_{kl} = \langle \log f, \psi_{kl}\rangle = \left\langle \log \frac{f}{2^{k/2}f_{kl}}, \psi_{kl}\right\rangle$, combining the Taylor expansion above yields

$$\widetilde{\log f}_{kl} = \langle \log f, \psi_{kl}\rangle = \frac{1}{2^{k/2}f_{kl}}\widetilde{f}_{kl} + \left\langle \frac{1}{2a^2}\left(\frac{f}{2^{k/2}f_{kl}} - 1\right)^2, \psi_{kl}\right\rangle$$

which implies

$$\left|\widetilde{\log f}_{kl}\right| \le C\left|\widetilde{f}_{kl}\right| + C2^k \int_{I_{kl}} \left(f - 2^{k/2}f_{kl}\right)^2.$$

We need to show there is a constant M' such that $\|\log f\|_{2,2}^{\alpha} \le M'$ and $\|\log f\|_{6,6}^{\alpha} \le M'$ for all $f \in \Sigma$. But we will only show the case $\|\log f\|_{6,6}^{\alpha} \le M'$, since the derivation for the other case $\|\log f\|_{2,2}^{\alpha} \le M'$ is similar and simpler. Define

$$\delta_{i,j,k,l} = \begin{cases} 1, & \text{if } I_{ij} \subset I_{kl} \\ 0, & \text{otherwise} \end{cases}$$

Let $c > 0$. Since $\sum_j \delta_{i,j,k,l} = 2^{i-k}$ for $i \geq k$, using Hölder inequality twice yields

$$\begin{aligned}
&\left(\int_{I_{kl}} \left(f - \overline{f}_k\right)^2\right)^m \\
= \quad & \left(\sum_{i=k}^{\infty} \sum_{j=1}^{2^i} \delta_{i,j,k,l} \widetilde{f}_{ij}^2\right)^m \\
\leq \quad & \left(\sum_{i=k}^{\infty} 2^{-ci} 2^{ci} \left(2^{i-k}\right)^{1-1/m} \left(\sum_{j=1}^{2^i} \delta_{i,j,k,l} \widetilde{f}_{ij}^{2m}\right)^{1/m}\right)^m \\
\leq \quad & 2^{-k(m-1)} \left(\sum_{i=k}^{\infty} 2^{-ci \cdot m/(m-1)}\right)^{m-1} \left(\sum_{i=k}^{\infty} 2^{i(m-1+cm)} \sum_{j=1}^{2^i} \delta_{i,j,k,l} \widetilde{f}_{ij}^{2m}\right) \\
\leq \quad & \frac{2^{-(cm+m-1)k}}{\left(1 - 2^{-cm/(m-1)}\right)^{m-1}} \left(\sum_{i=k}^{\infty} 2^{i(m-1+cm)} \sum_{j=1}^{2^i} \delta_{i,j,k,l} \widetilde{f}_{ij}^{2m}\right)
\end{aligned}$$

Then for $m = 6$ we write

$$\left(\int_{I_{kl}} \left(f - \overline{f}_k\right)^2\right)^6 \leq \frac{2^{-(6c+5)k}}{\left(1 - 2^{-6c/5}\right)^5} \left(\sum_{i=k}^{\infty} 2^{i(5+6c)} \sum_{j=1}^{2^i} \delta_{i,j,k,l} \widetilde{f}_{ij}^{12}\right).$$

Note that

$$\begin{aligned}
\sum_{k=0}^{\infty} 2^{k(6\alpha+2)} \sum_{l=1}^{2^k} \left|\widetilde{\log f}_{kl}\right|^6 \quad \leq \quad & C \sum_{k=0}^{\infty} 2^{k(6\alpha+2)} \sum_{l=1}^{2^k} \left|\widetilde{f}_{kl}\right|^6 \\
& + C \sum_{k=0}^{\infty} \sum_{l=1}^{2^k} 2^{k(6\alpha+3)} \left(\int_{I_{kl}} \left(f - J_{kl}(f)\right)^2\right)^6.
\end{aligned}$$

Since $f \in B_{6,6}^{\alpha}(M)$, we know that

$$\sum_{k=0}^{\infty} 2^{k(6\alpha+2)} \sum_{l=1}^{2^k} \left|\widetilde{f}_{kl}\right|^6$$

is uniformly bounded. For the second term we have

$$\begin{aligned}
&\sum_{k=0}^{\infty}\sum_{l=1}^{2^k} 2^{k(6\alpha+3)}\left(\int_{I_{kl}}\left(f-J_{kl}\left(f\right)\right)^2\right)^6 \\
\leq\;& \sum_{k=0}^{\infty} 2^{k(6\alpha+3)}\frac{2^{-(6c+5)k}}{\left(1-2^{-6c/5}\right)^5}\left(\sum_{i=k}^{\infty}2^{i(5+6c)}\sum_{j=1}^{2^i}\sum_{l=1}^{2^k}\delta_{i,j,k,l}\widetilde{f}_{ij}^{12}\right) \\
=\;& \sum_{k=0}^{\infty} 2^{k(6\alpha+3)}\frac{2^{-(6c+5)k}}{\left(1-2^{-6c/5}\right)^5}\left(\sum_{i=k}^{\infty}2^{i(5+6c)}\sum_{j=1}^{2^i}\widetilde{f}_{ij}^{12}\right) \\
=\;& \sum_{i=0}^{\infty}\sum_{k=0}^{i}\left(2^{k(6\alpha+3)}\frac{2^{-(6c+5)k}}{\left(1-2^{-6c/5}\right)^5}\right)2^{i(5+6c)}\sum_{j=1}^{2^i}\widetilde{f}_{ij}^{12}.
\end{aligned}$$

Let $c=\alpha$. Then

$$\sum_{k=0}^{i}\left(2^{k(6\alpha+3)}\frac{2^{-(6c+5)k}}{\left(1-2^{-6c/5}\right)^5}\right)\leq C,$$

and

$$2^{i(5+6c)}\widetilde{f}_{ij}^{12}=2^{i(6\alpha+2)}\widetilde{f}_{ij}^{6}\cdot 2^{3i}\widetilde{f}_{ij}^{6}\leq C2^{3i}\widetilde{f}_{ij}^{6}$$

due to the assumption $f\in B_{6,6}^{\alpha}\left(M\right)$, so we have

$$\sum_{kl}2^{k(6\alpha+3)}\left(\int_{I_{kl}}\left(f-J_{kl}\left(f\right)\right)^2\right)^6\leq C\sum_{i=0}^{\infty}2^{3i}\sum_{j=1}^{2^i}\widetilde{f}_{ij}^{6},$$

which is uniformly bounded again from the assumption $f\in B_{6,6}^{\alpha}\left(M\right)$. Thus $\left\|\log f\right\|_{6,6}^{\alpha}$ is uniformly bounded for all $f\in\Sigma$.

Lemma 2.9 *Suppose that P_i, Q_i are probability measures on a measurable space $\left(\Omega_i,\mathcal{F}_i\right)$, for $i=1,2,...,k$. Then*

$$H^2\left(\bigotimes_{i=1}^{k}P_i,\bigotimes_{i=1}^{k}Q_i\right)\leq\sum_{i=1}^{k}H^2\left(P_i,Q_i\right).$$

Lemma 2.10 *Suppose that ξ_j are i.i.d. $N\left(0,1\right)$, $j=1,\ldots,n$. The following two experiments given by observations*

$$\begin{aligned}
E_{1,n} &: z(j)=\log J_{j,n}\left(f\right)+\log\xi_j^2 \\
E_{1,n}^{1} &: z(j)=J_{j,n}\left(\log f\right)+\log\xi_j^2,
\end{aligned}$$

with unknown $f \in \Sigma$ are asymptotically equivalent.

Proof: We know $\log \xi_i^2$ has density

$$p(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\exp(t)}{2}\right) \exp\left(\frac{t}{2}\right).$$

We define

$$\begin{aligned} p_{1,j}(t) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\exp(t - \log J_{j,n}(f))}{2}\right) \exp\left(\frac{t - \log J_{j,n}(f)}{2}\right) \\ p_{2,j}(t) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\exp(t - J_{j,n}(\log f))}{2}\right) \exp\left(\frac{t - J_{j,n}(\log f)}{2}\right). \end{aligned}$$

Making use the general relation of Hellinger to Δ-distance and lemma 2.9 we obtain

$$\Delta^2\left(E_{1,n}, E_{1,n}^1\right) \leq \sum_{j=1}^{n} \int \left(p_{1,j}^{1/2} - p_{2,j}^{1/2}\right)^2 dt.$$

Obviously

$$dp^{1/2}(t)/dt = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{\exp(t)}{2}\right) \exp\left(\frac{t}{2}\right)(1 - \exp(t))$$

is uniformly bounded, then

$$\sum_{j=1}^{n} \int \left(p_{1,j}^{1/2} - p_{2,j}^{1/2}\right)^2 dt \leq C \sum_{j=1}^{n} (\log J_{j,n}(f) - J_{j,n}(\log f))^2.$$

¿From the lemma 2.8 we have

$$|\log J_{j,n}(f) - J_{j,n}(\log f)| \leq C J_{j,n}\left((f - J_{j,n}(f))^2\right).$$

Thus we have

$$\begin{aligned} \Delta^2\left(E_{1,n}, E_{1,n}^1\right) &\leq C \sum_{j=1}^{n} \left(J_{j,n}\left((f - J_{j,n}(f))^2\right)\right)^2 \\ &\leq C \sum_{j=1}^{n} J_{j,n}\left((f - J_{j,n}(f))^2\right) \\ &= C \sum_{k \geq \log n} \sum_{l=1}^{2^k} n \widetilde{f}_{kl}^2 \end{aligned}$$

which converges uniformly to 0, because Σ is compact in $B_{2,2}^{1/2}$.

Lemma 2.11 *Suppose that ξ_j are i.i.d. $N(0,1)$, $j = 1, \ldots, n$. The following two experiments given by observations*

$$\begin{aligned} E_{1,n} &: \quad z(j) = J_{j,n}(\log f/f_0) + \log \xi_j^2 \\ E_{1,n}^1 &: \quad z(j) = \log(1 + J_{j,n}(\log f/f_0)) + \log \xi_j^2, \end{aligned}$$

with unknown $f \in \Sigma_n(f_0)$ are asymptotically equivalent.

Proof: The proof is similar to that of Lemma 2.10. Applying Jensen's inequality yields

$$\sum_{i=1}^{n} (J_{j,n}(\log f_0/f))^4 \le Cn \int (\log f_0/f)^4 \to 0$$

Taylor expansion, combining with $|J_{j,n}(\log f_0/f)| \to 0$ implied by the equation above, gives

$$|J_{j,n}(\log f/f_0) + \log(1 + J_{j,n}(\log f_0/f))| \le C(J_{j,n}(\log f_0/f))^2,$$

then we have

$$\Delta^2\left(E_{1,n}, E_{1,n}^1\right) \le C \sum_{i=1}^{n} (J_{j,n}(\log f_0/f))^4 \to 0.$$

Lemma 2.12 *$E_{2,n}$ and $E_{2,n}^1$ are asymptotically equivalent, as $n \to \infty$.*

Proof: Similar to Brown and Low (1996), for some constant $C > 0$, we have

$$\begin{aligned} \Delta^2\left(E_{2,n}, E_{2,n}^1\right) &\le C \int \left(\log f - \sum_{j=1}^{n} J_{j,n}(\log f) I_{[(j-1)/n, j/n)}\right)^2 \\ &= C \sum_{k \ge \log n} \sum_{l=1}^{2^k} n \left(\widetilde{\log f}_{kl}\right)^2 \end{aligned}$$

where the right side converges uniformly to 0, because $\Sigma' = \{\log f : f \in \Sigma\}$ is compact in $B_{2,2}^{1/2}$ from the lemma 2.8.

Lemma 2.13 *We have*

$$\begin{aligned}\sum_{l=1}^{2^{k_0}} 2^{k_0} g_{k_0 l}^2 &\leq Cn^{1/2} \int |f-f_0|^2 \ = O\left((\log n)^{-4}\right) \\ \sum_{l=1}^{2^{k_0}} 2^{3k_0} g_{k_0 l}^4 &\leq Cn \int |f-f_0|^4 = O\left((\log n)^{-8}\right) \\ \sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} 2^k n^2 g_{kl}^6 &\leq Cn^{3/2} \int |f-f_0|^6 = O\left((\log n)^{-12}\right)\end{aligned}$$

and

$$\sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} 2^{3k} \ \widetilde{g}_{kl}^4 \leq \frac{1}{2} \sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} \left(2^k \ \widetilde{g}_{kl}^2 + 2^{5k} \ \widetilde{g}_{kl}^6\right) = O\left(n^{-(\alpha-1/2)}\right).$$

uniformly over $f \in \Sigma_n(f_0)$.

Proof: We only show the last two inequalities. By Jensen's inequality we have

$$\sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} 2^k n^2 g_{kl}^6 \leq C \sum_{k=k_0}^{k_1-1} \frac{n^2}{2^k} \int |g|^6$$

Since $|g| \leq C|f-f_0|$, then

$$\sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} 2^k n^2 g_{kl}^6 \leq C \frac{n^2}{2^{k_0}} \int |f-f_0|^6 = n^{3/2} \int |f-f_0|^6$$

which is $O\left(\log^6 n\right)$ from the equation (2.8).

The simple inequality $ab \leq \left(a^2+b^2\right)/2$ implies

$$\sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} 2^{3k} \ \widetilde{g}_{kl}^4 = \frac{1}{2} \sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} 2^{k/2} \ \widetilde{g}_{kl}^4 \cdot 2^{5k/2} \ \widetilde{g}_{kl}^3 \leq \frac{1}{2} \sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} \left(2^k \ \widetilde{g}_{kl}^2 + 2^{5k} \ \widetilde{g}_{kl}^6\right).$$

Lemma 2.8 tells us $g \in B_{2,2}^{\alpha}(M') \cap B_{6,6}^{\alpha}(M')$, for some $M' > 0$. From the definition of Besov norms in (2.4), we know

$$\sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} \left(2^{2\alpha k} \ \widetilde{g}_{kl}^2 + 2^{(6\alpha+2)k} \ \widetilde{g}_{kl}^6\right)$$

is uniformly bounded, which implies

$$\sum_{k=k_0}^{k_1-1} \sum_{l=1}^{2^k} \left(2^k \ \widetilde{g}_{kl}^2 + 2^{5k} \ \widetilde{g}_{kl}^6\right) = O\left(2^{-(2\alpha-1)k_0}\right) = O\left(n^{-(\alpha-1/2)}\right).$$

Lemma 2.14 *Let ξ be a real valued r.v. such that $E\xi = 0$, then we have*

$$\log E\exp(\xi) \leq \left(E\xi^4\right)^{1/2} E\exp(2|\xi|).$$

Proof: Let $\mu(t) = E\exp(t\xi)$ and $\psi(t) = \log\mu(t)$. Then, for some $0 \leq v \leq 1$,

$$\psi(1) = \psi(0) + \psi'(0) + \psi''(v).$$

Note that $\psi(0) = 0$, $\psi'(0) = 0$, $\mu(t) \geq 1$ by Jensen's inequality. We have, for any $0 \leq s \leq 1$,

$$\begin{aligned} 0 &\leq \psi''(s) = \mu(s)^{-2}\left\{E\xi^2\exp(s\xi) - (E\xi\exp(s\xi))^2\right\} \\ &\leq E\xi^2\exp(s\xi) \leq \left(E\xi^4\right)^{1/2}(E\exp(2s\xi))^{1/2} \leq \left(E\xi^4\right)^{1/2}(E\exp(2|\xi|))^{1/2}, \end{aligned}$$

which implies the result.

The following lemma was proved in Grama and Nussbaum (2003).

Lemma 2.15 *Let ξ be a real valued r.v. such that $E\xi = 0$, $0 < E\xi^2 < \infty$. Assume that Sakhanenko's condition*

$$\lambda E|\xi|^3\exp(\lambda|\xi|) \leq E\xi^2$$

holds for some $\lambda > 0$. Then for all $|t| \leq \lambda/3$ we have

$$E\exp(t\xi) \leq \exp\left(t^2 E\xi^2\right).$$

The lemma above implies the following lemma.

Lemma 2.16 *Suppose that $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d. random variables with mean 0 and finite variance. Let ξ_1 satisfies the Sakhanenko's condition above. Then for $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ and all $|t| \leq \lambda/3$ we have*

$$E\exp(tS_n) \leq \exp\left(t^2 ES_n^2\right),$$

and for $0 \le t < \lambda/3$ *we have*

$$P\left(|S_n| > x\right) \le 2\exp\left(t^2 ES_n^2 - tx\right),$$

which implies

$$P\left(|S_n| > x\right) \le 2\exp\left(-\frac{x^2}{4ES_n^2}\right), \text{ for } x < 2\lambda ES_n^2/3,$$

and for $0 \le c < \lambda/6$ *we have*

$$E\exp\left(c\left|S_n\right|\right) \le 1 + 2\exp\left(4c^2 ES_n^2\right).$$

Proof: ¿From lemma 2.15, we have

$$E\exp\left(tS_n\right) = \prod_{i=1}^{n} E\exp\left(t\xi_i\right) \le \prod_{i=1}^{n} \exp\left(t^2 E\xi_i^2\right) = \exp\left(t^2 ES_n^2\right).$$

Denote

$$F(x) = P\left(|S_n| > x\right).$$

Then for $x > 0$ and $0 \le t < \lambda/3$ we have

$$P\left(S_n > x\right)\exp\left(tx\right) \le E\exp\left(tS_n\right) \le \exp\left(t^2 ES_n^2\right),$$

and

$$P\left(S_n < -x\right)\exp\left(tx\right) \le E\exp\left(-tS_n\right) \le \exp\left(t^2 ES_n^2\right),$$

which implies

$$P\left(|S_n| > x\right) \le 2\exp\left(t^2 ES_n^2 - tx\right).$$

When $x \le 2\lambda ES_n^2/3$, the minimum of $\exp\left(t^2 ES_n^2 - tx\right)$ is achieved at $t = x/\left(2ES_n^2\right) < \lambda/3$.

Let $2c = t$. Integrating by parts we obtain

$$\begin{aligned} E\exp\left(c\left|S_n\right|\right) &= 1 + \int_0^\infty F(x)\, c\exp\left(cx\right) dx \\ &\le 1 + 2c\exp\left(t^2 ES_n^2\right)\int_0^\infty \exp\left(cx - tx\right) dx \\ &= 1 + 2\exp\left(4c^2 ES_n^2\right). \end{aligned}$$

Lemma 2.17 *Let $X_{k,l}$ be defined in (2.19), and let E_{P_2} be defined in the lemma (2.2). For $f \in \Sigma_n(f_0)$, we have*

$$\begin{aligned} E_{P_2}\left(\frac{X_{k,l}}{n/2^k}-1\right)^4 \left(n/2^k\right)^2 &\leq C\left(1+n^2 g_{kl}^4\right), \\ E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\right)^4 &\leq C. \end{aligned}$$

Proof: (i) Since

$$X_{k,l} = X_{k+1,2l} + X_{k+1,2l-1},$$

i.e.,

$$X_{k+1,2l-1} = \left(X_{k,l} - \left(X_{k+1,2l} - X_{k+1,2l-1}\right)\right)/2$$

, from the definition in equations (2.19) and (2.20) we may write

$$\zeta_{(k+1,2l-1)} = \frac{1}{2}\left(\zeta_{(kl)} + \widetilde{\zeta}_{(kl)}\right), \ \zeta_{(k+1,2l)} = \frac{1}{2}\left(\zeta_{(kl)} - \widetilde{\zeta}_{(kl)}\right).$$

Let $D_{k,l}$ be the set of all indices $\left(k', l'\right)$ with $k' > k$ such that $\left(2^{-k'}(l-1), 2^{-k'}l\right] \cap \left(2^{-k}(l-1), 2^{-k}l\right] \neq \phi$, then

$$\zeta_{(kl)} = \sum_{\left(k',l'\right)\in D_{k,l}} a_{k'l'} \frac{2^{k'}}{2^k} \tilde{\zeta}_{\left(k'l'\right)} + \frac{2^{k'}}{2^{k_0}} a_{k_0,l}\zeta_{(k_0,l)}$$

where $a_{k'l'}$ and $a_{k_0,l}$ are either 1 or -1, and we define

$$\eta_{(kl)} = \sum_{\left(k',l'\right)\in D_{k,l}} a_{k'l'} \frac{2^{k'}}{2^k} \widetilde{\eta}_{\left(k'l'\right)} + \frac{2^{k_0}}{2^k} a_{k_0,l}\eta_{(k_0,l)},$$

Then we have

$$\begin{aligned} E_{P_2}\left(\frac{X_{k,l}}{n/2^k}-1\right)^4 \left(n/2^k\right)^2 &= \frac{1}{4} E_{P_2}\left(\frac{\zeta_{(kl)}}{n/2^k}\right)^4 \left(n/2^k\right)^2 \\ &\leq 2E_{P_2}\left(\zeta_{(kl)} - \eta_{(kl)}\right)^4 \left(n/2^k\right)^{-2} + 2E_{P_2}\left(\frac{\eta_{(kl)}}{\sqrt{n/2^k}}\right)^4. \end{aligned}$$

Obviously $\eta_{(kl)}/\sqrt{n/2^k}$ is a normal random variable with mean $\sqrt{n}\widetilde{g}_{kl}$ and variance $n/2^k$, so we have

$$E_{P_2}\left(\frac{\eta_{(kl)}}{\sqrt{n/2^k}}\right)^4 \leq C\left(1+\left(\sqrt{n}\widetilde{g}_{kl}\right)^4\right)$$

by direct calculation.

If $n/2^k \geq n^{1/4}$, the Cauchy-Shwartz inequality gives

$$\begin{aligned}
& E_{P_2}\left(\zeta_{(kl)}-\eta_{(kl)}\right)^4\left(n/2^k\right)^{-2} \\
\leq\ & (\log n)^2 E_{P_2}\left(\sum_{k'=k_0}^{k-1}\left(\frac{2^{k'}}{2^k}\right)^4\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4+\left(\frac{2^{k'}}{2^k}\right)^4\left(\zeta_{k_0,l}-\eta_{k_0,l}\right)^4\right)\left(n/2^k\right)^{-2} \\
\leq\ & (\log n)^2 E_{P_2}\left(\sum_{k'=k_0}^{k-1}\left(n/2^{k'}\right)^{-2}\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4+\left(n/2^k\right)^{-2}\left(\zeta_{k_0,l}-\eta_{k_0,l}\right)^4\right)\left(n/2^k\right)^{-2}
\end{aligned}$$

¿From the equation (2.22) we have

$$\begin{aligned}
E_{P_2}\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4 &= E_{P_2}\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4\left(1\left(\widetilde{G}_{kl}\right)+1\left(\widetilde{G}_{kl}^c\right)\right) \\
&\leq CE_{P_2}\left(1+\frac{\left|\tilde{\eta}_{(k'l')}\right|}{\left(n/2^k\right)^{1/4}}+\frac{\left|\tilde{\eta}_{(k'l')}\right|^2}{n/2^{k'}}\right)^2+E_{P_2}\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4 1\left(\widetilde{G}_{kl}^c\right) \\
&\leq C\sqrt{n/2^{k'}}E_{P_2}\left(1+\frac{\left|\tilde{\eta}_{(k'l')}\right|^2}{n/2^{k'}}\right)^2+E_{P_2}\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4 1\left(\widetilde{G}_{kl}^c\right) \\
&\leq C\sqrt{n/2^{k'}}\left(1+\left(\sqrt{n}\widetilde{g}_{k'l'}\right)^4\right),
\end{aligned}$$

and

$$E_{P_2}\left(\zeta_{k_0,l}-\eta_{k_0,l}\right)^4 \leq C\left(1+\left(\sqrt{n}g_{k_0,l}\right)^4\right),$$

then we have

$$\begin{aligned}
\left(n/2^{k'}\right)^{-3/2}E_{P_2}\left(\tilde{\zeta}_{k'l'}-\widetilde{\eta}_{k'l'}\right)^4 &\leq C_1\left(1+2^{k'}n\widetilde{g}_{k'l'}^4\right)\leq C_2 \\
\left(n/2^k\right)^{-1}E_{P_2}\left(\zeta_{k_0,l}-\eta_{k_0,l}\right)^4 &\leq C_1\left(1+2^k n g_{k_0,l}^4\right)\leq C_2,
\end{aligned}$$

where the last step of boundedness is from lemma (2.13). So we have

$$E_{P_2}\left(\zeta_{(kl)}-\eta_{(kl)}\right)^4\left(n/2^k\right)^{-2} \leq C_1\sum_k\left(n/2^k\right)^{-1/2} \leq C_2.$$

If $n/2^k \le n^{1/4}$, a basic inequality $(a_1 + a_2 + \ldots + a_m)^4 \le 8a_1^4 + 8^2 a_2^4 + \ldots 8^k a_m^4$ implies

$$\begin{aligned}
& E_{P_2}\left(\zeta_{(kl)} - \eta_{(kl)}\right)^4 \left(n/2^k\right)^{-2} \\
\le\ & E_{P_2}\left(\sum_{k'=k_0}^{k-1} \frac{2^{k'}}{2^k}\left(\tilde{\zeta}_{k'l'} - \widetilde{\eta}_{k'l'}\right)^4 + \frac{2^{k'}}{2^k}\left(\zeta_{k_0,l} - \eta_{k_0,l}\right)^4\right)\left(n/2^k\right)^{-2} \\
\le\ & E_{P_2}\left(\sum_{k'=k_0}^{k-1} \left(n/2^k\right)^{-1}\left(\tilde{\zeta}_{k'l'} - \widetilde{\eta}_{k'l'}\right)^4 + \left(n/2^k\right)^{-1}\left(\zeta_{k_0,l} - \eta_{k_0,l}\right)^4\right).
\end{aligned}$$

Then we have

$$\begin{aligned}
\left(n/2^k\right)^{-1} E_{P_2}\left(\tilde{\zeta}_{k'l'} - \widetilde{\eta}_{k'l'}\right)^4 &\le C\left(\left(n/2^{k'}\right)^{-1/2} + \left(n/2^k\right)^{-1/2} n^2 \widetilde{g}_{k'l'}^4\right) \\
&\le C\left(\left(n/2^{k'}\right)^{-1/2} + 2^{3k'}\widetilde{g}_{k'l'}^4\right),
\end{aligned}$$

and

$$\left(n/2^k\right)^{-1} E_{P_2}\left(\zeta_{k_0,l} - \eta_{k_0,l}\right)^4 \le C_1\left(1 + 2^k n g_{k_0,l}^4\right)$$

so we have

$$E_{P_2}\left(\zeta_{(kl)} - \eta_{(kl)}\right)^4\left(n/2^k\right)^{-2} \le C\sum_{k'}\left(n/2^{k'}\right)^{-1/2} + C\sum_{k'} 2^{3k'}\widetilde{g}_{k'l'}^4 + C2^k n g_{k_0,l}^4,$$

which is bounded from lemma (2.13).

(ii) We have

$$\begin{aligned}
E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\right)^4 &= E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\right)^4\left(1\left(\frac{X_{k,l}}{n/2^k} < 2\right) + 1\left(\frac{X_{k,l}}{n/2^k} \ge 2\right)\right) \\
&\le 16 + E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\right)^4 1\left(\frac{X_{k,l}}{n/2^k} \ge 2\right).
\end{aligned}$$

Note that

$$\begin{aligned}
& E_{P_2}\left(\frac{X_{k,l}}{n/2^k}\right)^4 1\left(\frac{X_{k,l}}{n/2^k} \ge 2\right) \\
\le\ & \left(E\left(\frac{X_{k,l}}{n/2^k}\right)^{12}\right)^{1/2}\left(E1\left(\frac{X_{k,l}}{n/2^k} \ge 2\right)\right)^{1/4}\left(E\exp\left(4\sqrt{n}g_{kl}\frac{\eta_{(kl)}}{\sqrt{n/2^k}}\right)\right)^{1/4}
\end{aligned}$$

is bounded, since lemma (2.16) implies

$$E1\left(\frac{X_{k,l}}{n/2^k} \geq 2\right) = \exp\left(-C_1 n/2^k\right)$$

and direct calculation gives

$$E\exp\left(4\sqrt{n}g_{kl}\frac{\eta_{(kl)}}{\sqrt{n/2^k}}\right) = \exp\left(C_2 n g_{kl}^2\right),$$

and $ng_{kl}^2 \ll n/2^k$ from lemma (2.13).

BIBLIOGRAPHY

[1] Bohnenblust, H. F., Shapley, L.S. and Sherman, S. (1949), Reconnaissance in game theory. *RAND Research Memorandum* **RM-208** 118.

[2] Brown, L. D. , Cai, T., Low, M. G. & Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30** 688-707.

[3] Brown, L.D., Carter, A. V., Low M. G., and Zhang, C.-H. (2004) Asymptotic equivalence theory for density estimation, Poisson processes, and Gaussian with noise with drift. *Ann. Statist.* To appear.

[4] Brown, L. D. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384-2398.

[5] Carter, A. V. (2002). Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.* **30** (3) 708-730.

[6] Donoho, D. L. and Johnstone, I. (1995), Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Stat. Assoc.* **90**, 1200–1224.

[7] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion) *J. Roy. Statist. Soc. B* **57** 301-369.

[8] Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theor. Related Fields* **111** 493-533.

[9] Grama, I. and Nussbaum, M. (2003a). Asymptotic equivalence for nonparametric regression. *Math. Methods Statist.* **11** (1) 1-36.

[10] Grama, I. and Nussbaum, M. (2002b). A functional Hungarian construction for sums of independent random variables. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **38** (6) pp. 923-957.

[11] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df. I *Z. Wahrsch. verw. Gebiete* **32** 111-131.

[12] Le Cam, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35**,1419-1455.

[13] Le Cam, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3** xiv-xx. Univ. California Press, Berkeley.

[14] Le Cam, L., and Yang, G.L. (2000). *Asymptotics in Statistics.* Some Basic Concepts, 2nd ed. Springer, New York.

[15] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise, *Ann. Statist.* **24** 2399-2430.

[16] Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[17] Tsybakov, A. B. (1997). Asymptotically efficient signal estimation in L_2 under general loss functions. *Problems of Information Transmission* **33** no. 1, 78-88

[18] Van der Vaart, A. W.(2002). The statistical work of Lucien Le Cam. *Ann. Statist.* **30** (3) 631-682.

[19] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426-482.

[20] Carter, A. V. and Pollard, D. (2004). Tusnády's inequality revisited. Available at http://www.stat.yale.edu/˜pollard. To appear Ann. Statist.