9-2006

Uniform Convergence Rate of the SNP Density Estimator and Testing for Similarity of Two Unknown Densities

Kyoo-il KIM
Singapore Management University, kyookim@smu.edu.sg

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Uniform Convergence Rate of the SNP Density Estimator and Testing for Similarity of Two Unknown Densities

Kyoo il Kim

September 2006

Paper No. 20-2006

Uniform Convergence Rate of the SNP Density Estimator and Testing for Similarity of Two Unknown Densities

Kyoo il Kim*
School of Economics and Social Sciences
Singapore Management University

September 13, 2006

Abstract

This paper studies the uniform convergence rate of the turncated SNP (semi-nonparametric) density estimator. Using the uniform convergence rate result we obtain, we propose a test statistic testing the equivalence of two unknown densities where two densities are estimated using the SNP estimator and supports of densities are possibly unbounded.

Keywords: SNP Density Estimator, Uniform Convergence Rate, Comparison of Two Densities
JEL Classification: C12 C14 C16

1 Introduction

Gallant and Nychka (1987) introduce the semi-nonparametric (SNP) maximum likelihood estimation and establish its consistency. The SNP estimation is a convenient method that simultaneously estimates the parameters of a nonlinear model and the nonparametric density of a latent process by the quasi-maximum likelihood. Partly motivated by its computational convenience and its wide applicability compared to other nonparametric estimation, the SNP estimator has been popularly used although its convergence rate and the asymptotic distributional theory of this estimator are not known well. The distributional theory of the SNP is less of interest because the representation of the SNP estimator is parametric at any instance since it is a truncation estimator. Also it is noted that ignoring the truncation and treating the SNP estimator as fully parametric often provides reasonably accurate tests and confidence intervals (see Eastwood and Gallant (1991) and Fan, Zhang, and Zhang (2001)). Thus, the determination of a desirable truncation point has been more challenging (see Coppejans and Gallant (2002)).

However, with exceptions of Fenton and Gallant (1996a, b) that establish the L_1 convergence rate and Coppejans and Gallant (2002) that derive the convergence rate under the Hellinger metric, the convergence

*I would like to thank a co-editor and two anonymous referees for their correcting my errors and valuable suggestions, which much improved this paper. This paper was written during my doctoral study in UCLA. I am truly grateful to my advisor, Jinyong Hahn for his advice and encouragement throughout my life in UCLA. Correspondence: Kyoo il Kim, School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903, Tel: +65 6828–0876, Fax: +65 6828-0833, E-mail: kyookim@smu.edu.sg.

rate of the SNP density estimator has not been studied well. We often find that in deriving the asymptotic distribution of test statistics related with estimated densities, a convergence rate result in a stronger norm such as a uniform rate is often required. This paper establishes the uniform convergence rate of the SNP density estimator with a truncated compact support.

A referee of this paper points out that our proposed estimator does not achieve the optimal convergence rate obtained by Stone (1990) for log-spline models. However, we note that our approach with truncated Hermite series is still useful for several reasons. First, it makes easy to test the departure from the normality by construction. The insignificance of additional terms other than the first term in the series will suggest the normality (truncated normal). On the contrary, such test may require additional semiparametric testing procedure for the log-spline models. Second, our estimator is computationally easier than the Stone (1990)'s estimator. For example, $c(\theta)$ in Stone (1990) is hard to evaluate analytically for log-spline models based on quadratic and higher-order splines. Third, the proposed SNP density estimator can handle the case of unknown support or unbounded support by estimating the truncated version of such densities. Fourth, the approach proposed in this paper can extend to the case of data dependent support (for example, setting a support from the minimum value of observations to the maximum value of observations). Fifth, as noted above, some studies suggest that treating the SNP estimator as fully parametric ones (ignoring the truncation of the series) can still provide reasonably accurate inferences. Sixth, the SNP estimation is a convenient method that allows simultaneous estimations of the parameters of a nonlinear model and the nonparametric density as proposed in Gallant and Nychka (1987). Lastly, comparing two densities based on two SNP density estimators is feasible as illustrated in Section 3 of this paper.

Comparison of two densities has been considered in many studies since it can be used for comparing two samples or testing the independence assumption. However, in most of cases, only kernel density estimators are considered since the popular U-statistics approach is available for kernel estimators (see Bickel and Rosenblatt (1973), Rosenblatt (1975), Hall (1984), Robinson (1991), Li (1996), and Hong and White (2000)). This paper illustrates that the testing for similarity of two densities can be also implemented using the SNP density estimators. From the advantages of the SNP density estimator over the kernel estimator documented in literature (see Fenton and Gallant (1996b) for example), it is expected that this alternative testing procedure can entertain those advantages.

The organization of this paper is as follows. Section 2 establishes the convergence rate of the SNP estimator. A test statistic testing the similarity of two unknown densities using the SNP estimators is proposed in Section 3. Concluding Remarks follow in Section 4. Some technical details are presented in Appendix.

2 SNP Density Estimator with Truncated Support

We focus on the univariate case to simplify the notation since our objective here is to illustrate how we can obtain the uniform convergence rate, though the SNP estimation can be used for the multivariate densities. Extension to the multivariate case is not difficult. Now let a univariate random variable X follows a distribution with its density function $f_0(X)$. Hereafter the upper case X denotes a random variable and the lower case x denotes its realization. Then the SNP estimator of $f_0(X)$ for an i.i.d sample $\{x_i\}_{i=1}^n$ is defined as the quasi-maximum likelihood estimator of

$$\widehat{f}_{SNP} = \underset{f \in \mathcal{F}_n, \mu, \sigma > 0}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{1}{\sigma} f\left(\frac{x_i - \mu}{\sigma} \right) \right) \tag{1}$$

where

$$\mathcal{F}_n = \left\{ f : f(x, \xi) = \left(\sum_{j=1}^{K(n)} \xi_j x^{j-1} \right)^2 e^{-x^2/2} + \epsilon_0 \phi(x), \xi \in \Xi_n \right\},$$

where $\Xi_n = \left\{ \xi = (\xi_1, \ldots, \xi_{K(n)}) : \int_{-\infty}^{\infty} f_n(x, \xi) dx = 1 \right\}$, $K(n) \to \infty$ as $n \to \infty$ but $\frac{K(n)}{n} \to 0$, ϵ_0 is a small positive number, and $\phi(\cdot)$ denotes the standard normal density. It is worthwhile to note some structural aspects of the SNP density estimator. First, the leading term of the SNP expansion is the normal density with the weight function $e^{-x^2/2}$. This enables us to test the normality of the true density by testing whether $K(n) = 1$ against a general alternative $K(n) > 1$. Other weight functions as well as corresponding alternatives of the term $\epsilon_0 \phi(x)$ considered in Gallant and Nychka (1987) can generate different leading terms, although no applications has been found outside the normal case. Second, the term $\epsilon_0 \phi(x)$ serves as a lower bound of the density that insures $\int_{-\infty}^{\infty} (\ln f) f_0 dx$ exists for any f in $\mathcal{F}_n$ and prevents $\ln f$ from going out of range in optimizations. Lastly, the SNP estimator is positive and invariant to location and scale.

It is convenient to rewrite the class $\mathcal{F}_n$ in terms of Hermite polynomials:

$$\mathcal{F}_n = \left\{ f : f(x, \theta) = \left(\sum_{j=1}^{K(n)} \vartheta_j H_j(x) \right)^2 + \epsilon_0 \phi(x), \theta \in \Theta_n \right\}, \tag{2}$$

where $\Theta_n = \left\{ \theta = (\vartheta_1, \ldots, \vartheta_{K(n)}) : \sum_{j=1}^{K(n)} \vartheta_j^2 + \epsilon_0 = 1 \right\}$, $\{H_j(t)\}$ are defined recursively as

$$\begin{aligned} H_1(t) &= (\sqrt{2\pi})^{-1/2} e^{-t^2/4}, H_2(t) = (\sqrt{2\pi})^{-1/2} t e^{-t^2/4}, \\ H_j(t) &= [t H_{j-1}(t) - \sqrt{j-1} H_{j-2}(t)] / \sqrt{j}, \text{ for } j \geq 3. \end{aligned} \tag{3}$$

These Hermite polynomials are orthonormal $\int_{-\infty}^{\infty} H_j^2(z) dx = 1$, $\int_{-\infty}^{\infty} H_j(z) H_k(z) dx = 0$, $j \neq k$ and bounded $|H_j| < \widetilde{H} = 0.6862127$ (see Abramowitz and Stegun (1972, Chapter 22)). Under the following assumption that requires f_0 has exponential tails:

Assumption 2.1 (Fenton and Gallant (1996a)) *(i) The observed data $\{x_i\}$ are a random sample from the continuous density f_0. For some $\kappa \geq 0$, f_0 can be written as $f_0(x) = h_{f_0}^2(x) e^{-x^2/2} + \epsilon_0 \phi(x)$ where $\int_{-\infty}^{\infty} \left[\frac{d^j}{dx^j} h_{f_0}(x) \right]^2 e^{-x^2/2} dx < \infty$, for $j = 0, 1, \ldots, \kappa$. (ii) For every $a_1 > 0$ and a_0 in $(-\infty, \infty)$ there exists b_0 and $b_1 > 0$ such that* $\Pr\left(x^2 > a_0 + a_1 B\right) \leq b_0 e^{-b_1 \sqrt{B}}$,

Fenton and Gallant (1996a) establishes the L_1 convergence of the SNP estimator as

$$\int_{-\infty}^{\infty} \left| \widehat{f}_{SNP}(x) - f_0(x) \right| dx = o_s\left(n^{-1/2+\alpha/2+\delta}\right) + o\left(n^{-\kappa\alpha}\right) \text{ a.s.}$$

where the SNP is truncated at $K(n) = O(n^{\alpha})$, $\alpha > 0$ and δ is arbitrary small positive constant. Similarly the convergence rate under the Hellinger metric is derived in Coppejans and Gallant (2002) with $K(n) =$

$n^{(1-\overline{\delta})/(\kappa+1)}$ for given $\overline{\delta} \in (0,1)$,

$$d_H(\widehat{f}_{SNP}, f_0) \equiv \left\{ \int_{-\infty}^{\infty} \left(\widehat{f}_{SNP}^{1/2}(x) - f_0^{1/2}(x) \right)^2 dx \right\}^{1/2} = n^{-\kappa(1-\overline{\delta})/(2\kappa+2)} \text{ a.s.}$$

where $d_H(\cdot,\cdot)$ denotes the Hellinger distance under Assumption 2.1 (i) and restricting f_0 to have a largest mode and to satisfy the tail condition $\lim_{K\to\infty} \sup_{-\infty<x<\infty} \frac{h_{f_0}^2(x)+\left(\sum_{j=1}^K \vartheta_j H_j(x)\right)^2}{\left(\sum_{j=1}^K \vartheta_j H_j(x)\right)^2+\epsilon_0/\sqrt{2\pi}} \le C$ for some $C < \infty$.

Differently from the previous literature of the SNP estimation, in this paper we restrict the support of a density to be truncated as a compact subset $\mathcal{X}$ of $\mathcal{R}$. One can interpret our estimator as a truncated distribution version of the SNP density estimator. Noting the boundary problem of a nonparametric density estimation, our approach can be accepted as a valid one in the sense that we focus on an estimator that shows a good performance on the support of interest. One could work on the range of a compact support that increases as the sample size gets larger but this requires additional technicalities that are beyond the scope of this paper. In any case the SNP estimation under the compact support may be of interest by itself. Interestingly, we find that there exists a one-to-one mapping between the coefficients of the SNP density estimator with unbounded support and the coefficients of the SNP density estimator with a truncated support, which we propose as follows (see Appendix A).

Now let 1_A be the indicator function of the set A and consider the class $\underline{\mathcal{F}}_n$ in terms of truncated Hermite polynomials defined on $\mathcal{X}$ such that

$$\underline{\mathcal{F}}_n = \left\{ f : f(x,\theta) = \left(\sum_{j=1}^{K(n)} \vartheta_j w_{jK}(x) \right)^2 + \epsilon_0 \frac{\phi(x) 1_{\mathcal{X}}(x)}{\int_{\mathcal{X}} \phi(x)dx}, \theta \in \Theta_n \right\}, \tag{4}$$

where $\Theta_n = \left\{ \theta = (\vartheta_1, \ldots, \vartheta_{K(n)}) : \sum_{j=1}^{K(n)} \vartheta_j^2 + \epsilon_0 = 1 \right\}$ and a triangular array $\{w_{jK}(x)\}$ are defined below. First we define $\overline{w}_{jK}(x) = \frac{H_j(x)1_{\mathcal{X}}(x)}{\sqrt{\int_{\mathcal{X}} H_j^2(x)dx}}$that is bounded by

$$\sup_{x\in\mathcal{X}, j\le K} |\overline{w}_{jK}(x)| \le \frac{1}{\sqrt{\min_{j\le K} \int_{\mathcal{X}} H_j^2(x)dx}} \sup_{x\in\mathcal{X}} |H_j(x)| < C\widetilde{H}$$

for some constant $C < \infty$, since $\int_{\mathcal{X}} H_j^2(x)dx$ is bounded away from zero for all j and $|H_j(x)| < \widetilde{H}$ uniformly over x and j. Denoting $\overline{W}^K(x) = (\overline{w}_{1K}(x), \ldots, \overline{w}_{KK}(x))'$, further define $Q_{\overline{W}} = \int_{\mathcal{X}} \overline{W}^K(x) \overline{W}^K(x)' dx$ and its symmetric matrix square root as $Q_{\overline{W}}^{-1/2}$. Now let

$$W^K(x) \equiv (w_{1K}(x), \ldots, w_{KK}(x))' \equiv Q_{\overline{W}}^{-1/2} \overline{W}^K(x) \tag{5}$$

then by construction, we have $\int_{\mathcal{X}} W^K(x) W^K(x)' = I_K$. Then these truncated and transformed Hermite polynomials are orthonormal $\int_{\mathcal{X}} w_{jK}^2(x)dx = 1$, $\int_{\mathcal{X}} w_{jK}(x) w_{kK}(x)dx = 0, j \neq k$ from which the condition $\sum_{j=1}^{K(n)} \vartheta_j^2 + \epsilon_0 = 1$ follows since for any f in $\underline{\mathcal{F}}_n$, we have $\int_{\mathcal{X}} f dx = 1$. Now define $\zeta(K) = \sup_{x\in\mathcal{X}} \left\| W^K(x) \right\|$ using a matrix norm $\|A\| = \sqrt{\text{tr}(A'A)}$ for a matrix A, which is the Euclidian norm for a vector. Then, we can obtain $\zeta(K) = O(\sqrt{K})$ as shown in Lemma B.1. If the range of $\mathcal{X}$ is sufficiently large, then $\int_{\mathcal{X}} H_j^2(x)dx \approx 1$ and $Q_{\overline{W}} \approx I_K$ and hence $w_{jK} \approx H_j$ which implies immediately

$$\sup_{x\in\mathcal{X}} \left\| W^K(x) \right\| \approx \sup_{x\in\mathcal{X}} \sqrt{\sum_{j=1}^K H_j^2} \le \sqrt{K\widetilde{H}^2} = O\left(\sqrt{K}\right).$$

For brevity, we set $(\mu, \sigma) = (0, 1)$ without loss of generality (wlog) hereafter noting the SNP estimator is invariant to location and scale. The SNP density estimator with the compact support is obtained by solving

$$\widehat{f} = \underset{f \in \underline{\mathcal{F}}_n}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \ln f(x_i), \;\; f \in \underline{\mathcal{F}}_n \tag{6}$$

or equivalently

$$\widehat{f} = f(x, \widehat{\theta}_K), \, \widehat{\theta}_K = \underset{\theta \in \Theta_n}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \ln f(x_i, \theta)$$

for $f \in \underline{\mathcal{F}}_n$ defined in (4)[1]. Note this truncated version of the SNP density estimator is positive and invariant to location and scale. The term $\epsilon_0 \phi(x) 1_{\mathcal{X}}(x) / \int_{\mathcal{X}} \phi(x) dx$ helps $\ln f$ not to go out of range in optimizations so that the excess influence of one or several summands when $f(\cdot)$ are arbitrary small is avoided. Interestingly, we find there exists a one-to-one relationship between elements of $\mathcal{F}_n$ and $\underline{\mathcal{F}}_n$ (see Appendix A). In this sense, we interpret our version of the SNP estimator as a truncated version of the original SNP estimator (with unbounded support) developed by Gallant and his co-authors.

2.1 Uniform Convergence Rate

To derive the uniform convergence rate, we impose

Assumption 2.2 *(i) The observed data $\{x_i\}$ are an i.i.d. sample from the continuous density f_0, (ii) $f_0(x)$ is s-times continuously differentiable with $s \geq 3$, (iii) uniformly bounded from above and bounded away from zero on its compact support $\mathcal{X}$, (iv) $f_0(x)$ has the form of $f_0(x) = h_{f_0}^2(x) e^{-x^2/2} + \epsilon_0 \frac{\phi(x) 1_{\mathcal{X}}(x)}{\int_{\mathcal{X}} \phi(x) dx}$ for arbitrary small positive number ϵ_0.*

Note that differently from Fenton and Gallant (1996a) and Coppejans and Gallant (2002), we do not require a tail condition since we impose the compact support condition. Under Assumption 2.2, we obtain

Theorem 2.1 *Suppose Assumption 2.2 holds and $\frac{\zeta(K)^2 K}{n} \to 0$. Then, for $K = O(n^{\alpha})$ with $\alpha < \frac{1}{3}$, we have*

$$\sup_{x \in \mathcal{X}} \left| \widehat{f}(x) - f_0(x) \right| = O\left(\zeta(K)^2\right) o_p\left(n^{-1/2 + \alpha/2 + \delta}\right) + O\left(\zeta(K)^2 K^{-s/2}\right) \tag{7}$$

for arbitrary small positive constant δ.

Compared to the L_1 convergence rate derived by Fenton and Gallant (1996a), Theorem 2.1 shows that the uniform convergence rate is slower than the L_1 rate since it requires uniformity over the support of X. The derived uniform convergence rate depends on the choice of sieve. For example, if someone uses the power series sieve rather than the Hermite polynomials (truncated and transformed as in (5)), he/she will achieve a slower convergence rate since $\zeta(K) = O(K)$ for the power series sieve compared to $\zeta(K) = O(\sqrt{K})$ for the Hermite polynomials. When we choose K such that two terms in (7) are balanced, which is achieved when $\alpha = 1/(1+s)$, the convergence rate will be $n^{-(s-2)/2(1+s)}$ noting $o_p\left(n^{-1/2+\alpha/2+\delta}\right) = O_p\left(\sqrt{\frac{K}{n}}\right)$ effectively. Thus, our rate cannot attain Stone's (1982) bound on the best obtainable rate, which equals to $\left(n^{-1} \ln n\right)^{s/(1+2s)}$. In the next section, we derive Theorem 2.1.

[1] To implement this estimation, in practice one needs to pick the optimal length of series according to some criteria, since the approximation precision depends on the choice of smoothing parameter K. One could use the Coppejans and Gallant (2002)'s method, which is a cross-validation strategy based on the integrated squared error (ISE). *Leave-one-out* or *Leave-one partition-out* (so called *hold-out-sample cross-validation*) method can be adopted depending on the size and property of data.

2.2 Derivation of the Convergence Rate

Though we have a particular sieve in mind for the SNP estimator, we derive the convergence rate result for a general sieve that satisfies some conditions. First note, according to Theorem 8, p.90, in Lorentz (1986), we can approximate a v-times continuously differentiable function h such that there exists a K-vector γ_K that satisfies

$$\sup_{z\in\mathcal{Z}}\left|h(z)-R^K(z)'\gamma_K\right| = O(K^{-\frac{v}{\dim(z)}}) \tag{8}$$

where $\mathcal{Z}$ is the compact support of h and $R^K(z)$ is a triangular array of polynomials. Now let $f_0(x) = h_{f_0}^2(x)e^{-x^2/2} + \epsilon_0\frac{\phi(x)}{\int_{\mathcal{X}}\phi(x)dx}$ and assume $h_{f_0}(x)$ (and hence $f_0(x)$) is s-times continuously differentiable on $\mathcal{X}$. Denote a K-vector $\theta_K = (\vartheta_{1K},\ldots,\vartheta_{KK})'$. Then, there exists a θ_K such that

$$\sup_{x\in\mathcal{X}}\left|h_{f_0}(x)-e^{x^2/4}W^K(x)'\theta_K\right| = O(K^{-s}) \tag{9}$$

by (8) noting $h_{f_0}(x)$ is s-times continuously differentiable over $\mathcal{X}$, $\mathcal{X}$ is compact, and $\left\{e^{x^2/4}w_{jK}(x)\right\}$ are linear combinations of power series. (9) implies that

$$\sup_{x\in\mathcal{X}}\left|h_{f_0}(x)e^{-x^2/4}-W^K(x)'\theta_K\right| \le \sup_{x\in\mathcal{X}}e^{-x^2/4}\sup_{x\in\mathcal{X}}\left|h_{f_0}(x)-e^{x^2/4}W^K(x)'\theta_K\right| = O(K^{-s}) \tag{10}$$

since $\sup\limits_{x\in\mathcal{R}} e^{-x^2/4}\le 1$. From this result, now it is shown below that $\sup_{x\in\mathcal{X}}|f_0(x)-f(x,\theta_K)| = O\left(\zeta(K)K^{-s}\right)$. First, note (10) implies

$$W^K(x)'\theta_K - O(K^{-s}) \le h_{f_0}(x)e^{-x^2/4} \le W^K(x)'\theta_K + O(K^{-s})$$

from which it follows that

$$\begin{aligned}\left(W^K(x)'\theta_K - O(K^{-s})\right)^2 - \left(W^K(x)'\theta_K\right)^2 &\le h_{f_0}^2(x)e^{-x^2/2} - \left(W^K(x)'\theta_K\right)^2 \\ &\le \left(W^K(x)'\theta_K + O(K^{-s})\right)^2 - \left(W^K(x)'\theta_K\right)^2\end{aligned} \tag{11}$$

assuming $W^K(x)'\theta_K$ is positive and $O(K^{-s})$ is a positive sequence without loss of generality. Now, note that

$$\sup_{x\in\mathcal{X}}\left|W^K(x)'\theta\right| \le \sup_{x\in\mathcal{X}}\left\|W^K(x)\right\|\|\theta\| = O\left(\zeta(K)\right) \tag{12}$$

by the Cauchy-Schwarz inequality and from $\|\theta\|^2 < 1$ for any $\theta\in\Theta_n$ by construction. Now applying the mean value theorem to the upper bound of (11), we have

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\left|\left(W^K(x)'\theta_K + O(K^{-s})\right)^2 - \left(W^K(x)'\theta_K\right)^2\right| = \sup_{x\in\mathcal{X}}\left|\left(2W^K(x)'\theta_K + O(K^{-s})\right)O(K^{-s})\right| \\ \le\ &\sup_{x\in\mathcal{X}}\left|2W^K(x)'\theta_K\right|O(K^{-s}) + O(K^{-2s}) = O(\zeta(K)K^{-s})\end{aligned}$$

where the last result is from (12). Similarly for the lower bound, we have

$$\sup_{x\in\mathcal{X}}\left|\left(W^K(x)'\theta_K - O(K^{-s})\right)^2 - \left(W^K(x)'\theta_K\right)^2\right| = O(\zeta(K)K^{-s}).$$

From (11), it follows that $\sup_{x\in\mathcal{X}}\left|h_{f_0}^2(x)e^{-x^2/2}-\left(W^K(x)'\theta_K\right)^2\right|=O(\zeta(K)K^{-s})$ and hence

$$\sup_{x\in\mathcal{X}}|f_0(x)-f(x,\theta_K)|=O\left(\zeta(K)K^{-s}\right). \tag{13}$$

Now to establish the convergence rate of the SNP estimator, a pseudo true density function is introduced where the pseudo true density is given by

$$f_K^*(x)=\left(W^K(x)'\theta_K^*\right)^2+\epsilon_0\frac{\phi(x)}{\int_{\mathcal{X}}\phi(x)dx}$$

such that

$$f_K^*(x)=f(x,\theta_K^*)\text{ and }\theta_K^*=\operatorname*{argmax}_{\theta}E\left[\ln f\left(X,\theta\right)\right]\text{ for }f\in\underline{\mathcal{F}}_n. \tag{14}$$

We first obtain

Lemma 2.1 *Suppose Assumption 2.2 holds. Then for θ_K in (9) and θ_K^* defined in (14), we have $\|\theta_K-\theta_K^*\|=O\left(K^{-s/2}\right)$ and thus*

$$\sup_{x\in\mathcal{X}}|f_0(x)-f_K^*(x)|=O\left(\zeta(K)^2K^{-s/2}\right). \tag{15}$$

Proofs of lemmas and technical derivations are in Appendix. Lemma 2.1 establishes the distance between the true density and the pseudo true density. Now the stochastic order of $\left\|\widehat{\theta}_K-\theta_K^*\right\|$ is derived using the uniform law of large numbers. Define $\widehat{Q}_n(\theta)=\frac{1}{n}\sum_{i=1}^n\ln f(x_i,\theta)$ and $Q(\theta)=E\left[\ln f\left(x,\theta\right)\right]$. Then we have

$$\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta)-Q(\theta)\right|=o_p\left(n^{-1/2+\alpha/2+\delta}\right) \tag{16}$$

for all sufficiently small $\delta>0$ from Lemma D.5. For $\|\theta-\theta^o\|\le o(\eta_n)$, we also have

$$\sup_{\|\theta-\theta^o\|\le o(\eta_n),\theta\in\Theta_n}\left|\widehat{Q}_n(\theta)-\widehat{Q}_n(\theta^o)-\left(Q(\theta)-Q(\theta^o)\right)\right|=o_p\left(\eta_n n^{-1/2+\alpha/2+\delta}\right) \tag{17}$$

as shown in Lemma D.6. From (16) and (17), it follows that

Lemma 2.2 *Suppose Assumption 2.2 holds and $\frac{\zeta(K)^2K}{n}\to 0$. Then for $K(n)=O\left(n^{\alpha}\right)$ with $\alpha<\frac{1}{3}$,*

$$\left\|\widehat{\theta}_K-\theta_K^*\right\|=o_p\left(n^{-1/2+\alpha/2+\delta}\right)$$

where δ is an arbitrary small positive constant.

Thus we obtain

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|\widehat{f}(x)-f_K^*(x)\right| &= \sup_{x\in\mathcal{X}}\left|\left(W^K(x)'(\widehat{\theta}_K-\theta_K^*)\right)\left(W^K(x)'(\widehat{\theta}_K+\theta_K^*)\right)\right| \\ &\le C_1\left(\sup_{x\in\mathcal{X}}\left\|W^K(x)\right\|\right)^2\left\|\widehat{\theta}_K-\theta_K^*\right\|=O\left(\zeta(K)^2\right)o_p\left(n^{-1/2+\alpha/2+\delta}\right)\end{aligned} \tag{18}$$

since $\|\theta\|^2<1$ for any $\theta\in\Theta_n$. Finally, we obtain Theorem 2.1 as

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|\widehat{f}(x)-f_0(x)\right| &\le \sup_{x\in\mathcal{X}}\left|\widehat{f}(x)-f_K^*(x)\right|+\sup_{x\in\mathcal{X}}|f_K^*(x)-f_0(x)| \\ &= O\left(\zeta(K)^2\right)o_p\left(n^{-1/2+\alpha/2+\delta}\right)+O\left(\zeta(K)^2K^{-s/2}\right)\end{aligned} \tag{19}$$

from (15) and (18).

3 Comparison of Densities Using a Symmetric Kullback-Leibler Divergence Measure

Here[2] we are interested in testing the equivalence of two densities where these densities are estimated using two different samples. Examples include the comparison of income distribution across two regions, groups, or time. Suppose that we are testing

$$H_0 : f = g \tag{20}$$

against

$$H_A : f \neq g.$$

Testing the symmetry of $f(x)$ also fits into this framework where we put $g(x) = f(-x)$.

A natural measure to compare $f(\cdot)$ and $g(\cdot)$ will be the integrated squared error given by $I_s(f(z),g(z)) = \int_{\mathcal{Z}}(f(z) - g(z))^2 dz$ assuming f and g have the common compact support $\mathcal{Z}$. Under the null we have $I_s(f(z),g(z)) = 0$. Li (1996) develops a test statistic of this sort when both densities are estimated using a kernel method. Other possible measures for the distance of two density functions are the Kullback-Leibler (KL) information distance or the Hellinger metric. The KL measure is entertained in Ullah and Singh (1989), Robinson (1991), and Hong and White (2000) when they eventually test the affinity of two densities. Ullah and Singh (1989) and Robinson (1991) are based on a first order theory while Hong and White (2000) improves on these delivering higher power by using a second order theory. Su and White (2003) use a class of the Hellinger metric when they test for the conditional independence restriction by comparing a joint density and the product of two marginal densities. The KL measure is defined by $I_{KL} = \int_{\mathcal{Z}} (\ln f(z) - \ln g(z)) f(z)dz$ or $I_{KL} = \int_{\mathcal{Z}} (\ln g(z) - \ln f(z))\, g(z)\, dz$ which are equally zero under the null and have positive values under the alternative as shown in Kullback and Leibler (1951). However, it is noted that the KL information distance is not a proper distance measure, since it is not symmetric although it still serves as a valid discrepancy measure. Here we propose a variation of the Kullback-Leibler measure which is symmetric and nonnegative. We define

$$I(f, g) = \int_{\mathcal{Z}} (\ln f(z) - \ln g(z))\, f(z)dz + \int_{\mathcal{Z}} (\ln g(z) - \ln f(z))\, g(z)\, dz \tag{21}$$

which has zero value under the null but is strictly positive under the alternative by construction. It is also symmetric, $I(f, g) = I(g, f)$. The proposed test statistic will be constructed as a sample analogue of (21).

Now suppose that $\widehat{f}$ and $\widehat{g}$ are estimated using the samples $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$, respectively. The sizes of two samples do not have to be the same but here we impose the same size for notational convenience. For these two sets of data, we assume

Assumption 3.1 *Suppose $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ are iid, respectively and $x_i \perp y_j$ for all $i, j = 1, \ldots n$. x_i and y_i have the density function f and g, respectively, which are continuous and bounded away from zero on their common compact support $\mathcal{Z}$.*

Here we focus on the SNP density estimators, though the result presented here is quite general as in Robinson (1991). Different estimators require different primitive conditions and trimming devices that ensures the sufficient conditions for validity of the test statistic we propose here. We first present general conditions for consistency and asymptotic validity of the test statistic and show these conditions are satisfied

[2] In this section, we do not distinguish a random variable from its realizatoin to make notation simple unless otherwise noted.

for the SNP density estimator. A set of conditions in the case of the kernel density estimation can be found in Robinson (1991).

3.1 Test Statistic and Its Asymptotic Property

Here we derive a test statistic based a symmetric KL measure of (21) as

$$\begin{aligned} I(\widehat{f},\widehat{g}) &= \int_{\mathcal{Z}} (\ln \widehat{f}(z) - \ln \widehat{g}(z))\widehat{f}(z)dz + \int_{\mathcal{Z}} (\ln \widehat{g}(z) - \ln \widehat{f}(z))\widehat{g}(z)dz \\ &= \int_{\mathcal{Z}} (\ln \widehat{f}(z) - \ln \widehat{g}(z))d\widehat{F}(z) + \int_{\mathcal{Z}} (\ln \widehat{g}(z) - \ln \widehat{f}(z))d\widehat{G}(z) \end{aligned}$$

where $d\widehat{F}(z) = \widehat{f}(z)dz$ and $d\widehat{G}(z) = \widehat{g}(z)dz$. Noting $\int_{\mathcal{Z}}(\ln \widehat{f}(z) - \ln \widehat{g}(z))d\widehat{F}(z) \approx \frac{1}{n}\sum_{i=1}^{n}\left(\ln \widehat{f}(x_i) - \ln \widehat{g}(x_i)\right)$ and $\int_{\mathcal{Z}}(\ln \widehat{g}(z) - \ln \widehat{f}(z))d\widehat{G}(z) \approx \frac{1}{n}\sum_{i=1}^{n}\left(\ln \widehat{g}(y_i) - \ln \widehat{f}(y_i)\right)$, we propose a test statistic[3] in the following form

$$\widehat{I}(\widehat{f},\widehat{g}) = \frac{1}{n-1}\sum_{i\in\mathcal{N}_x}\left(\ln \widehat{f}(x_i) - \ln \widehat{g}(x_{i+1})\right) + \frac{1}{n-1}\sum_{i\in\mathcal{N}_y}\left(\ln \widehat{g}(y_i) - \ln \widehat{f}(y_{i+1})\right) \tag{22}$$

where $\mathcal{N}_x$ and $\mathcal{N}_y$ are subsets of $\{1,2,,\ldots,n-1\}$. $\mathcal{N}_x$ and $\mathcal{N}_y$ trim out those observations of $\widehat{f}(\cdot) \leq \delta_f(n)$ or $\widehat{g}(\cdot) \leq \delta_g(n)$ for chosen positive values of $\delta_f(n)$ and $\delta_g(n)$ that tend to zero as $n \to \infty$[4]. To be precise, we define $\mathcal{N}_x = \left\{i : 1 \leq i \leq n-1 \text{ such that } \widehat{f}(x_i) > \delta_f(n) \text{ and } \widehat{g}(x_{i+1}) > \delta_g(n)\right\}$ and $\mathcal{N}_y = \left\{i : 1 \leq i \leq n-1 \text{ such that } \widehat{f}(y_{i+1}) > \delta_f(n) \text{ and } \widehat{g}(y_i) > \delta_g(n)\right\}$.

However, unfortunately, $\sqrt{n}\widehat{I}(\widehat{f},\widehat{g})$ will have a degenerate distribution under the null similarly as discussed in Robinson (1991) and cannot be used as a reasonable statistics. To resolve this problem, we entertain a modification of (22) in spirit of Robinson (1991) as

$$\widehat{I}_\gamma(\widehat{f},\widehat{g}) = \frac{1}{n_\gamma-1}\sum_{i\in\mathcal{N}_x}c_i(\gamma)\left(\ln \widehat{f}(x_i) - \ln \widehat{g}(x_{i+1})\right) + \frac{1}{n_\gamma-1}\sum_{i\in\mathcal{N}_y}c_i(\gamma)\left(\ln \widehat{g}(y_i) - \ln \widehat{f}(y_{i+1})\right)$$

where for a nonnegative constant γ,

$$\begin{aligned} c_i(\gamma) &= 1+\gamma \quad \text{if } i \text{ is } \textit{odd} \\ &= 1-\gamma \quad \text{if } i \text{ is } \textit{even} \end{aligned}$$

and n_γ is defined as[5]

$$n_\gamma = n+\gamma \quad \text{if } n \text{ is } \textit{odd} \text{ and } n_\gamma = n \quad \text{if } n \text{ is } \textit{even.} \tag{23}$$

---

[3] By construction of the test statistic, the way to *name* $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ will affect the test statistic since $\ln \widehat{g}(x_{i+1})$ is not evaluated at the first observation of $\{x_i\}_{i=1}^n$ and $\ln \widehat{f}(x_i)$ is not evaluated at the last observation of $\{x_i\}_{i=1}^n$. Similarly $\ln \widehat{f}(y_{i+1})$ is not evaluated at the first observation of $\{y_i\}_{i=1}^n$ and $\ln \widehat{g}(y_i)$ is not evaluated at the last observation of $\{y_i\}_{i=1}^n$. However, if the sample size is large enough, then this *naming* effect will be negligible.

[4] This trimming is a usual device in an inference procedure for nonparametric estimations. Even though the SNP density estimator that we are interested in is always positive by construction differently from higher order Kernel estimators, we still introduce this trimming device to avoid the excess influence of one or several summands when $\widehat{f}(\cdot)$ or $\widehat{g}(\cdot)$ are arbitrary small.

[5] Consider $s(n) \equiv \frac{1}{n_\gamma}\sum_{i=1}^n c_i(\gamma)$ when $n = 2m$ and $n = 2m+1$, respectively. It follows that $s(2m) = \frac{((1+\gamma)m+(1-\gamma)m)}{n_\gamma} = \frac{2m}{n_\gamma} = \frac{n}{n_\gamma}$ and $s(2m+1) = \frac{1}{n_\gamma}((1+\gamma)m + (1-\gamma)m + (1+\gamma)) = \frac{2m+1+\gamma}{n_\gamma} = \frac{n+\gamma}{n_\gamma}$. Thus, by constructing n_r as (23), we have $s(n) = 1$.

We let $I = I(f,g)$, $\widetilde{I}_\gamma = \widehat{I}_\gamma(f,g)$, and $\widehat{I}_\gamma = \widehat{I}_\gamma(\widehat{f},\widehat{g})$ for notational simplicity. Now note, for any increasing sequence $d(n)$, any positive $C < \infty$, and γ, we have

$$\Pr(d(n)\widehat{I}_\gamma < C) \leq \Pr(d(n)\left|\widehat{I}_\gamma - I\right| > d(n)I - C) \leq \Pr(\left|\widehat{I}_\gamma - I\right| > I/2)$$

when n is sufficiently large and (20) is not true (i.e. $I > 0$). Since the probability $\Pr(\left|\widehat{I}_\gamma - I\right| > I/2)$ goes to zero under the alternative as long as $\widehat{I}_\gamma \underset{p}{\longrightarrow} I$, one could construct a test statistic of the form

$$\text{Reject (20) when } d(n)\widehat{I}_\gamma > C. \tag{24}$$

Therefore, as long as $\widehat{I}_\gamma \underset{p}{\longrightarrow} I$, (24) is a valid test consistent against all departures from (20). We call a test statistic consistent against one direction of departure from the null hypothesis if the rejection probability approaches one as the sample size gets large regardless of the size of that departure. The following lemma establishes the conditions for $\widehat{I}_\gamma \underset{p}{\longrightarrow} I$. First, we let $E_h[\cdot]$ denote an expectation operator that takes expectation with respect to a density h.

Lemma 3.1 *Suppose Assumption 3.1 holds. Suppose that (i) $E_f\left[|\ln f|\right] < \infty$ and $E_g\left[|\ln f|\right] < \infty$, (ii) $E_f\left[|\ln g|\right] < \infty$ and $E_g\left[|\ln g|\right] < \infty$, (iii) $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i \notin \mathcal{N}_x) = o(1)$ and $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i \notin \mathcal{N}_y) = o(1)$.*

Further suppose (iv)

$$\frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\left(\widehat{f}(x_i)/f(x_i)\right) \underset{p}{\longrightarrow} 0,\ \frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\left(\widehat{f}(y_{i+1})/f(y_{i+1})\right) \underset{p}{\longrightarrow} 0,$$

$$\frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\left(\widehat{g}(x_{i+1})/g(x_{i+1})\right) \underset{p}{\longrightarrow} 0,\ \text{and}\ \frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\left(\widehat{g}(y_i)/g(y_i)\right) \underset{p}{\longrightarrow} 0,$$

then we have $\widehat{I}_\gamma \underset{p}{\longrightarrow} I$.

See Appendix for the proof. The larger the order of $d(n)$ is, while $d(n)\widehat{I}_\gamma$ preserves the limiting normal distribution with zero mean under the null, the test statistic will have higher powers. In what follows, we show that we can achieve this with $d(n) = O\left(\sqrt{n}\right)$. Suppose

$$\begin{aligned}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\left(\widehat{f}(x_i)/f(x_i)\right) &= o_p\left(\sqrt{n}\right),\ \sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\left(\widehat{g}(x_{i+1})/g(x_{i+1})\right) = o_p\left(\sqrt{n}\right), \\ \sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\left(\widehat{g}(y_i)/g(y_i)\right) &= o_p\left(\sqrt{n}\right),\ \sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\left(\widehat{f}(y_{i+1})/f(y_{i+1})\right) = o_p\left(\sqrt{n}\right),\end{aligned} \tag{25}$$

then it follows immediately that

$$\widehat{I}_\gamma - \widetilde{I}_\gamma = o_p\left(\frac{1}{\sqrt{n}}\right)$$

for all $\gamma \geq 0$. This implies that the asymptotic distribution of $\sqrt{n}\widehat{I}_\gamma$ will be identical to that of $\sqrt{n}\widetilde{I}_\gamma$ under the null, which means the effect of nonparametric estimation is negligible. Now consider, under the null $f = g$,

$$\begin{aligned}
\widetilde{I}_\gamma &= \frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\left(\ln f(x_i) - \ln g(x_{i+1})\right) + \frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\left(\ln g(y_i) - \ln f(y_{i+1})\right) \qquad (26)\\
&= \frac{2\gamma}{n_\gamma - 1}\sum_{i\in\mathcal{Q}}\left(\ln f(x_{i+1}) - \ln f(x_i)\right) + \frac{1+\gamma}{n_\gamma - 1}\ln f(x_1) - \frac{c_{\max\mathcal{Q}+1}(\gamma)}{n_\gamma - 1}\ln f(x_{\max\mathcal{Q}+1})\\
&\quad + \frac{2\gamma}{n_\gamma - 1}\sum_{i\in\mathcal{Q}}\left(\ln g(y_{i+1}) - \ln g(y_i)\right) + \frac{1+\gamma}{n_\gamma - 1}\ln g(y_1) - \frac{c_{\max\mathcal{Q}+1}(\gamma)}{n_\gamma - 1}\ln g(y_{\max\mathcal{Q}+1})\\
&\quad + O_p\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x)\right) + O_p\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_y)\right)
\end{aligned}$$

where $\mathcal{Q} = \{i : 1 \leq i \leq n-1,\ i \text{ even}\}$. We are willing to choose $\gamma > 0$.[6] Now suppose (25) holds. Further suppose

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x) = o\left(\frac{1}{\sqrt{n}}\right),\ \frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_y) = o\left(\frac{1}{\sqrt{n}}\right) \qquad (27)$$

and assume

$$E\left[|\ln f(x_i)|^2\right] < \infty \text{ and } E\left[|\ln g(y_i)|^2\right] < \infty, \qquad (28)$$

then, under the null of (20), we have

$$\frac{1}{\sqrt{n/2}}\sum_{i\in\mathcal{Q}}\left(\ln f(x_{i+1}) - \ln f(x_i)\right) \underset{d}{\longrightarrow} N(0,\Sigma) \text{ and } \frac{1}{\sqrt{n/2}}\sum_{i\in\mathcal{Q}}\left(\ln g(y_{i+1}) - \ln g(y_i)\right) \underset{d}{\longrightarrow} N(0,\Sigma) \qquad (29)$$

where $\Sigma = E\left[\left(\ln f(x_{i+1}) - \ln f(x_i)\right)^2\right]$ or $E\left[\left(\ln g(y_{i+1}) - \ln g(y_i)\right)^2\right]$ that equals to $2\mathrm{Var}[\ln f(\cdot)]$ or $2\mathrm{Var}[\ln g(\cdot)]$ by the Lindberg-Levy central limit theorem under the null. Therefore from (26) and (29), we conclude that under (25), (27), and (28),

$$\sqrt{n}\widehat{I}_\gamma = \sqrt{n}\widetilde{I}_\gamma + o_p(1) \underset{d}{\longrightarrow} N(0, 4\gamma^2\Sigma) \qquad (30)$$

under (20) for any $\gamma > 0$ noting $f = g$ and $\{x_1, \ldots, x_n, y_1, \ldots y_n\}$ are iid under the null. Finally, we conclude that

$$\frac{\sqrt{n}\widehat{I}_\gamma}{2\gamma\widehat{\Sigma}^{\frac{1}{2}}} \underset{d}{\longrightarrow} N(0,1)$$

for any $\widehat{\Sigma} = \Sigma + o_p(1)$. Possible candidates of $\widehat{\Sigma}$ will be

$$\begin{aligned}
\widehat{\Sigma}_1 &= 2\left(\frac{1}{2n}\sum_{i=1}^{n}\left\{(\ln\widehat{f}(x_i))^2 + (\ln\widehat{f}(y_i))^2\right\} - \left\{\frac{1}{2n}\sum_{i=1}^{n}(\ln\widehat{f}(x_i) + \ln\widehat{f}(y_i))\right\}^2\right) \text{ or}\\
\widehat{\Sigma}_2 &= 2\left(\frac{1}{2n}\sum_{i=1}^{n}\left\{(\ln\widehat{g}(x_i))^2 + (\ln\widehat{g}(y_i))^2\right\} - \left\{\frac{1}{2n}\sum_{i=1}^{n}(\ln\widehat{g}(x_i) + \ln\widehat{g}(y_i))\right\}^2\right)
\end{aligned}$$

or its average $\widehat{\Sigma}_3 = \frac{\widehat{\Sigma}_1 + \widehat{\Sigma}_2}{2}$. All of these are consistent under (28) and under Condition (iii) and (iv) of Lemma 3.1 and (20). We summarize the result as follows

[6] It is obvious that the distribution of (30) degenerates when $\gamma = 0$. See Robinson (1991) for related discussion.

Theorem 3.1 *Suppose Assumption 3.1 holds. Provided that (25), (27), and (28) hold under (20), we have*

$$\widehat{\tau}_\gamma \equiv \frac{\sqrt{n}\widehat{I}_\gamma}{2\gamma\widehat{\Sigma}^{\frac{1}{2}}} \xrightarrow[d]{} N(0,1)$$

for any $\gamma > 0$ and thus we reject (20) if $\widehat{\tau}_\gamma > C_\alpha$ where C_α is the size α one sided critical value of the standard normal distribution.

3.2 Primitive Conditions for the SNP Estimator

In what follows, we show that all the conditions for Lemma 3.1 and Theorem 3.1 are satisfied for the SNP density estimator. Here we should note that Lemma 3.1 holds whether or not the null ($f = g$) is true while Theorem 3.1 is required to hold only under the null. We start with conditions for Lemma 3.1. First, note Condition (i) and (ii) in Lemma 3.1 immediately hold since f and g are assumed to be continuous and $\mathcal{Z}$ is compact. Condition (iii) of Lemma 3.1 is verified as follows. For $\delta_f(n)$ and $\delta_g(n)$ that are positive numbers tending to zero as $n \to \infty$, consider

$$\begin{aligned}
& \frac{1}{n-1}\sum_{i=1}^{n-1} \Pr(i \notin \mathcal{N}_x) \qquad (31)\\
\leq\ & \frac{1}{n-1}\sum_{i=1}^{n-1} \Pr\left(\widehat{f}(x_i) \leq \delta_f(n) \text{ or } \widehat{g}(x_{i+1}) \leq \delta_g(n)\right)\\
\leq\ & \frac{1}{n-1}\sum_{i=1}^{n-1} \Pr\left(\left|\widehat{f}(x_i) - f(x_i)\right| + \delta_f(n) \geq f(x_i) \text{ or } \left|\widehat{g}(x_{i+1}) - g(x_{i+1})\right| + \delta_g(n) \geq g(x_{i+1})\right)\\
\leq\ & \frac{1}{n-1}\sum_{i=1}^{n-1} \Pr\left(\sup_{x\in\mathcal{Z}}\left|\widehat{f}(x) - f(x)\right| + \delta_f(n) \geq f(x_i) \text{ or } \sup_{x\in\mathcal{Z}}\left|\widehat{g}(x) - g(x)\right| + \delta_g(n) \geq g(x_{i+1})\right)
\end{aligned}$$

and hence as long as $\sup_{x\in\mathcal{Z}}\left|\widehat{f}(x) - f(x)\right| = o_p(1)$, $\sup_{x\in\mathcal{Z}}\left|\widehat{g}(x) - g(x)\right| = o_p(1)$, $\delta_f(n) = o(1)$, and $\delta_g(n) = o(1)$, we have $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i \notin \mathcal{N}_x) = o_p(1)$ since f and g are bounded away from zero. Therefore under $\alpha \leq \frac{1}{3} - \frac{2}{3}\delta$ and $s > 2$,

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i \notin \mathcal{N}_x) = o_p(1)$$

from Theorem 2.1 and also we can show $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i \notin \mathcal{N}_y) = o_p(1)$ similarly.

Condition (iv) of Lemma 3.1 is easily established from the uniform convergence rate result. Using $|\ln(1+t)| \leq 2|t|$ in a neighborhood of $t = 0$, consider

$$\begin{aligned}
& \left|\frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\widehat{f}(y_{i+1})/f(y_{i+1})\right| \qquad (32)\\
\leq\ & (1+\gamma)\sup_{y\in\mathcal{Z}}\left|\ln\widehat{f}(y) - \ln f(y)\right| \leq (1+\gamma)\sup_{y\in\mathcal{Z}} 2\left|\frac{f(y) - \widehat{f}(y)}{\widehat{f}(y)}\right|\\
=\ & O\left(\zeta(K)^2\right) o_p\left(n^{-1/2+\alpha/2+\delta}\right) + O\left(\zeta(K)^2 K^{-s/2}\right)
\end{aligned}$$

from Theorem 2.1 and since $\widehat{f}(\cdot)$ is bounded away from zero. Therefore, $\frac{1}{n_\gamma-1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\frac{\widehat{f}(y_{i+1})}{f(y_{i+1})} = o_p(1)$ under $\alpha \leq \frac{1}{3}-\frac{2}{3}\delta$ and $s>2$ with $K=O(n^\alpha)$ noting $\zeta(K)=O\left(\sqrt{K}\right)$. Similarly we can show $\frac{1}{n_\gamma-1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\frac{\widehat{f}(x_i)}{f(x_i)} = o_p(1)$, $\frac{1}{n_\gamma-1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\frac{\widehat{g}(x_{i+1})}{g(x_{i+1})} = o_p(1)$, and $\frac{1}{n_\gamma-1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\frac{\widehat{g}(y_i)}{g(y_i)} = o_p(1)$ under $\alpha\leq\frac{1}{3}-\frac{2}{3}\delta$ and $s>2$.

Now we establish conditions for Theorem 3.1. Again (28) immediately holds since f and g are assumed to be continuous and $\mathcal{Z}$ is compact. Next, we show (27). From (31) and the Markov inequality, we have

$$\begin{aligned}
&\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x) &(33)\\
\leq\;& \sup_{x\in\mathcal{Z}}\left(\frac{1}{f(x)^\tau}\right)\frac{1}{n-1}\sum_{i=1}^{n-1}E\left[\left|\widehat{f}(x_i)-f(x_i)+\delta_f(n)\right|^\tau\right]\\
&+\sup_{x\in\mathcal{Z}}\left(\frac{1}{g(x)^\tau}\right)\frac{1}{n-1}\sum_{i=1}^{n-1}E\left[\left|\widehat{g}(x_{i+1})-g(x_{i+1})+\delta_g(n)\right|^\tau\right]
\end{aligned}$$

and hence $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x) = o_p(\frac{1}{\sqrt{n}})$ as long as

$$\sup_{x\in\mathcal{Z}}\left|\widehat{f}(x)-f(x)\right|^\tau = o_p(\frac{1}{\sqrt{n}}),\ \sup_{x\in\mathcal{Z}}\left|\widehat{g}(x)-g(x)\right|^\tau = o_p(\frac{1}{\sqrt{n}}),$$

$\delta_f(n)^\tau = o(\frac{1}{\sqrt{n}})$, and $\delta_g(n)^\tau = o(\frac{1}{\sqrt{n}})$ noting $f(\cdot)$ and $g(\cdot)$ are bounded away from zero. Note

$$\begin{aligned}
\sup_{x\in\mathcal{Z}}\left|\widehat{f}(x)-f(x)\right|^\tau &= o_p\left(n^{(-1/2+3\alpha/2+\delta)\tau}\right)+O\left(n^{(1-s/2)\alpha\tau}\right) \text{ and}\\
\sup_{x\in\mathcal{Z}}\left|\widehat{g}(x)-g(x)\right|^\tau &= o_p\left(n^{(-1/2+3\alpha/2+\delta)\tau}\right)+O\left(n^{(1-s/2)\alpha\tau}\right)
\end{aligned}$$

from Theorem 2.1 and $\zeta(K)=O\left(\sqrt{K}\right)$ letting $K=O\left(n^\alpha\right)$. In particular, we choose $\tau=4$ and hence $\sup_{x\in\mathcal{Z}}\left|\widehat{f}(x)-f(x)\right|^\tau = o_p(\frac{1}{\sqrt{n}})$ and $\sup_{x\in\mathcal{Z}}\left|\widehat{g}(x)-g(x)\right|^\tau = o_p(\frac{1}{\sqrt{n}})$ under $\alpha\leq\frac{1}{4}-\frac{2}{3}\delta$ and $s\geq 2+\frac{1}{4\alpha}$. $\delta_f(n)^\tau = o(\frac{1}{\sqrt{n}})$ and $\delta_g(n)^\tau = o(\frac{1}{\sqrt{n}})$ hold under $\delta_f(n)=o(n^{-\frac{1}{8}})$ and $\delta_g(n)=o(n^{-\frac{1}{8}})$. Therefore, under $\alpha\leq\frac{1}{4}-\frac{2}{3}\delta$, $s\geq 2+\frac{1}{4\alpha}$, $\delta_f(n)=o(n^{-\frac{1}{8}})$, and $\delta_g(n)=o(n^{-\frac{1}{8}})$, we have $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x) = o\left(\frac{1}{\sqrt{n}}\right)$ and similarly we can show $\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_y) = o\left(\frac{1}{\sqrt{n}}\right)$.

Next, we verify (25) as

Lemma 3.2 *Suppose Assumption 3.1 holds. Further suppose Assumption 2.2 holds for f and g and $\frac{\zeta(K)^2K}{n}\to 0$. Then,*

$$\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\left(\widehat{f}(x_i)/f(x_i)\right) = o_p\left(\sqrt{n}\right),\ \sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln\left(\widehat{g}(x_{i+1})/g(x_{i+1})\right) = o_p\left(\sqrt{n}\right),$$

$$\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\left(\widehat{g}(y_i)/g(y_i)\right) = o_p\left(\sqrt{n}\right),\ \text{and}\ \sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln\left(\widehat{f}(y_{i+1})/f(y_{i+1})\right) = o_p\left(\sqrt{n}\right)$$

for $K(n)=O(n^\alpha)$ with $\alpha<\frac{1}{4}$, $s\geq\max\{\frac{1}{2\alpha},2+\frac{1}{4\alpha}\}$, $\delta_f(n)=o(n^{-\frac{1}{8}})$, and $\delta_g(n)=o(n^{-\frac{1}{8}})$.

See Appendix for the proof. Thus, we have provided primitive conditions under which all the conditions of Lemma 3.1 and Theorem 3.1 are satisfied. We summarize the results in the following lemma.

Lemma 3.3 *Suppose Assumption 3.1 and Assumption 2.2 hold for f and g. Then, under $\alpha < \frac{1}{3}$ and $s \geq 3$, all the conditions for Lemma 3.1 hold. Also for $\alpha < \frac{1}{4}$ and $s \geq 4$, under the null of (20), all the conditions of Theorem 3.1 are satisfied for the SNP estimators when we use trimming devices, $\delta_f(n) = o(n^{-\frac{1}{8}})$ and $\delta_g(n) = o(n^{-\frac{1}{8}})$.*

4 Conclusion

This paper establishes the uniform convergence rate of the SNP density estimator with a compact support. The estimator studied in this paper can be interpreted as a truncated distribution version of the Gallant and Nychka (1987)'s SNP estimator. It turns out that the convergence rate depends on the choice of sieve and the (truncated and transformed) Hermite polynomials provide a preferred rate.

Based on this result, in the spirit of Robinson (1991), we are able to test the equivalence of two unknown densities where two densities are estimated by the SNP estimator. The proposed test statistic entertains a version of the Kullback-Leibler information distance, which is symmetric and nonnegative. Other than the computational advantages, the SNP density estimator is of interest for implementing the test, since there are certain cases where the kernel estimator cannot be used but still the SNP method can be used especially when the density of interest is derived from the data whose DGP (data generating process) is a functional of the density of interest. Examples include several auction models where the distribution of valuations is derived from the bidding data.

Appendix

A Relationship Between $\mathcal{F}_n$ and $\underline{\mathcal{F}}_n$

Consider a specification of the SNP density estimator with unbounded support as in $\mathcal{F}_n$,

$$f(x,\theta) = (H^{(J)}(x)'\theta_J)^2 + \epsilon_0\phi(x) \tag{34}$$

where $H^{(J)}(x) = (H_1(x), H_2(x), \ldots, H_J(x))$ and $H_j(x)$'s are Hermite polynomials constructed recursively. Now consider a truncated version of the density on a truncated compact support $\mathcal{X}$ based on (34),

$$\overline{f}(x,\theta) = \frac{(H^{(J)}(x)'\theta_J)^2 + \epsilon_0\phi(x)}{\int_{\mathcal{X}}(H^{(J)}(x)'\theta_J)^2dx + \epsilon_0\int_{\mathcal{X}}\phi(x)dx}.$$

Let B be a $J\times J$ matrix whose elements are b_{ij}'s where $b_{ii} = \sqrt{\int_{\mathcal{X}} H_i(x)^2dx}$ for $i = 1,\ldots,J$ and $b_{ij} = 0$ for $i \neq j$. Recall that $\overline{W}^J(x) = B^{-1}H^{(J)}(x)$, $Q_{\overline{W}} = \int_{\mathcal{X}}\overline{W}^J(x)\overline{W}^J(x)'dx$, and $W^J(x) = Q_{\overline{W}}^{-1/2}\overline{W}^J(x) = Q_{\overline{W}}^{-1/2}B^{-1}H^{(J)}(x)$. Using those notations, we obtain

$$\begin{aligned}
\overline{f}(x,\theta) &= \frac{(H^{(J)}(x)'\theta_J)^2 + \epsilon_0\phi(x)}{\int_{\mathcal{X}}(H^{(J)}(x)'\theta_J)^2dx + \epsilon_0\int_{\mathcal{X}}\phi(x)dx} = \frac{\theta_J' H^{(J)}(x)H^{(J)}(x)'\theta_J + \epsilon_0\phi(x)}{\theta_J'\int_{\mathcal{X}}H^{(J)}(x)H^{(J)}(x)'dx\theta_J + \epsilon_0\int_{\mathcal{X}}\phi(x)dx} \\
&= \frac{\theta_J' BQ_{\overline{W}}^{1/2}Q_{\overline{W}}^{-1/2}B^{-1}H^{(J)}(x)H^{(J)}(x)'B^{-1}Q_{\overline{W}}^{-1/2}Q_{\overline{W}}^{1/2}B\theta_J + \epsilon_0\phi(x)}{\theta_J' B\int_{\mathcal{X}}B^{-1}H^{(J)}(x)H^{(J)}(x)'B^{-1}dxB\theta_J + \epsilon_0\int_{\mathcal{X}}\phi(x)dx} \\
&= \frac{\theta_J' BQ_{\overline{W}}^{1/2}W^J(x)W^J(x)'Q_{\overline{W}}^{1/2}B\theta_J + \epsilon_0\phi(x)}{\theta_J' B\int_{\mathcal{X}}\overline{W}^J(x)\overline{W}^J(x)'dxB\theta_J + \epsilon_0\int_{\mathcal{X}}\phi(x)dx} = \frac{\theta_J' BQ_{\overline{W}}^{1/2}W^J(x)W^J(x)'Q_{\overline{W}}^{1/2}B\theta_J + \epsilon_0\phi(x)}{\theta_J' BQ_{\overline{W}}^{1/2}Q_{\overline{W}}^{1/2}B\theta_J + \epsilon_0\int_{\mathcal{X}}\phi(x)dx}.
\end{aligned}$$

Now let $\widetilde{\theta}_J = Q_{\overline{W}}^{1/2}B\theta_J$ and $\epsilon_0 = \widetilde{\epsilon}_0/\int_{\mathcal{X}}\phi(x)dx$. Then, we simplify[7]

$$\overline{f}(x,\theta) = \frac{\widetilde{\theta}_J' W^J(x)W^J(x)'\widetilde{\theta}_J + \widetilde{\epsilon}_0\phi(x)/\int_{\mathcal{X}}\phi(x)dx}{\widetilde{\theta}_J'\widetilde{\theta}_J + \widetilde{\epsilon}_0} = \left(W^J(x)'\widetilde{\theta}_J\right)^2 + \widetilde{\epsilon}_0\phi(x)/\int_{\mathcal{X}}\phi(x)dx \tag{35}$$

by restricting $\widetilde{\theta}_J'\widetilde{\theta}_J + \widetilde{\epsilon}_0 = 1$ such that $\widetilde{\theta}_J \in \Theta_n$. Note that (35) coincides with the specification we consider in $\underline{\mathcal{F}}_n$ of (4). This illustrates why the proposed SNP estimator is a truncated version of the original SNP estimator with unbounded support. It is also noted that the relationship between the (truncated) parameters of the original density and the (truncated) parameters of the proposed specification is explicit as $\widetilde{\theta}_J = Q_{\overline{W}}^{1/2}B\theta_J$.

B Bound of the Truncated Hermite Series

Throughout the appendix, we use the following notation. C, C_1, C_2, $\ldots$ denote generic positive constants. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalue of a matrix A, respectively. We use $\dim(q)$ to denote the dimension of a vector q. We apply the Triangle Inequality or the Cauchy-Schwarz Inequality without indicating them unless it is necessary.

[7] Throughout the Appendix, we will suppress the indicator function $1_{\mathcal{X}}(\cdot)$ denoting the truncation in the truncated density function to make notation simple.

Lemma B.1 *Suppose $W^K(x)$ is given by (5). Then,* $\sup_{x\in\mathcal{X}}\left\|W^K(x)\right\| = \zeta(K) = O(\sqrt{K})$.

Proof. First we show that $Q_{\overline{W}}$ is nonsingular for sufficiently large K. By taking the support of X symmetric around zero wlog[8], we have $\int_{\mathcal{X}}\overline{w}_{iK}(x)\overline{w}_{jK}(x)dx = 0$ whenever $i+j$ is odd. When $i+j$ is even and $i \neq j$, we have $\int_{\mathcal{X}}\overline{w}_{iK}(x)\overline{w}_{jK}(x)dx \neq 0$ but $\left|\int_{\mathcal{X}}\overline{w}_{iK}(x)\overline{w}_{jK}(x)dx\right| < 1$ by construction since $\left|\int_{\mathcal{X}} H_i(x)H_j(x)dx\right| < \left|\sqrt{\int_{\mathcal{X}} H_i(x)^2dx}\sqrt{\int_{\mathcal{X}} H_j(x)^2dx}\right|$ from the Cauchy-Schwarz inequality noting $\int_{\mathcal{X}} 1(H_i(x) \neq H_j(x))dx > 0$. Now note that the determinant of a matrix $A \equiv [a_{ij}]$ can be computed by writing down all permutations of $\{1, 2, \ldots, \dim(A)\}$, taking each permutation $(ijk\ldots)$ as the subscripts in $a_{1i}a_{2j}a_{3k}\ldots$ and summing with signs obtained by $\epsilon_p = (-1)^{i(p)}$, where $i(p)$ denotes number of permutation inversions in permutation p (Muir (1960), p.16), and ϵ_p is the permutation symbol. Using this method, a close examination of $Q_{\overline{W}}$ reveals that the determinant of $Q_{\overline{W}}$ with $\dim(Q_{\overline{W}}) = K$ is computed as for $L \geq 2$,

$$\det(Q_{\overline{W}}) = \begin{cases} K=2 & 1 \\ K=2L & 1+\sum_{l=1}^{L-1} a_l(K) - \sum_{l=1}^{L} b_l(K) \\ K=2L-1 & 1-b(K) \end{cases}$$

where $a_l(K)$ and $b_l(K)$ is $K-2$ products of positive values less than one and $b(K)$ is $K-1$ products of positive values less than one. Hence whenever K is odd, $\det(Q_{\overline{W}})$ is obviously greater than zero. Noting $\sum_{l=1}^{L} b_l(K) \leq L\cdot\left(\max_{l\leq L} b_l(K)\right) \leq \frac{K}{2}(d)^{K-2} \to 0$ for some $d<1$, we find that $\det(Q_{\overline{W}}) > 0$ for sufficiently large K when K is even. Thus we conclude that $Q_{\overline{W}}$ is nonsingular for sufficiently large K. Actually we can make $Q_{\overline{W}}$ is nonsingular always by simply taking K odd. From $Q_{\overline{W}}$'s positive-definiteness (positive-semidefinite by construction and nonsingular), it follows that $\lambda_{\min}(Q_{\overline{W}})$ is bounded away from zero at least for K odd or sufficiently large K. Finally note

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left\|W^K(x)\right\| &= \sup_{x\in\mathcal{X}}\sqrt{\overline{W}^K(x)'Q_{\overline{W}}^{-1}\overline{W}^K(x)} \leq \sup_{x\in\mathcal{X}}\sqrt{\lambda_{\min}^{-1}(Q_{\overline{W}})\,\overline{W}^K(x)'\overline{W}^K(x)} \\ &\leq \sqrt{\lambda_{\min}^{-1}(Q_{\overline{W}})\sup_{x\in\mathcal{X}}\sum_{j=1}^{K}\overline{w}_{jK}^2(x)} \leq \sqrt{\lambda_{\min}^{-1}(Q_{\overline{W}})\,K\left(\widetilde{H}\right)^2} = O\left(\sqrt{K}\right).\end{aligned}$$

■

C Proof of Lemma 2.1

Define $Q_0 = E[\ln f_0(x)]$ and $Q(\theta) = E[\ln f(x,\theta)]$. Then, by definition, $\theta_K^* = \underset{\theta}{\operatorname{argmax}}\, Q(\theta)$. Consider

$$\underset{\theta}{\operatorname{argmin}}\, Q_0 - Q(\theta) = \underset{\theta}{\operatorname{argmax}}\, Q(\theta)$$

which implies that among the parametric family $\{f(x,\theta) : \theta = (\vartheta_1, \ldots, \vartheta_K)\}$, $Q(\theta_K^*)$ will have the minimum distance to Q_0 noting for all $\theta \in \Theta_n$, $Q(\theta) \leq Q_0$ from the information inequality (see Gallant (1987, p.484)). First, we show that

$$Q_0 - Q(\theta_K) = O(K^{-s}). \tag{36}$$

[8] For any bounded support, we can make the support symmetric around zero by shifting the support.

Using $|\ln(1+t)| \leq 2|t|$ in a neighborhood of $t=0$, consider

$$\begin{aligned}
|Q_0 - Q(\theta_K)| &\leq E\left[\left|\ln \frac{f_0(X)}{f(X,\theta_K)}\right|\right] \leq E\left[2\left|\frac{f_0(X)}{f(X,\theta_K)} - 1\right|\right] \qquad (37)\\
&= 2\int_{\mathcal{X}} \frac{1}{f(x,\theta_K)} |f_0(x) - f(x,\theta_K)| f_0(x)dx \\
&= 2\int_{\mathcal{X}} \frac{f_0(x)}{f(x,\theta_K)} \left|h_{f_0}(x)e^{-x^2/4} + W^K(x)'\theta_K\right| \left|h_{f_0}(x)e^{-x^2/4} - W^K(x)'\theta_K\right| dx \\
&\leq 2\sup_{x\in\mathcal{X}} \frac{\sqrt{f_0(x)}}{f(x,\theta_K)} \sup_{x\in\mathcal{X}} \left|h_{f_0}(x)e^{-x^2/4} - W^K(x)'\theta_K\right| E\left[\frac{\left|h_{f_0}(X)e^{-X^2/4}\right| + \left|W^K(X)'\theta_K\right|}{\sqrt{f_0(X)}}\right]
\end{aligned}$$

and by the triangular inequality. Note

$$E\left[\frac{\left|h_{f_0}(X)e^{-x^2/4}\right|}{\sqrt{f_0(X)}}\right] \leq \sqrt{E\left[\frac{h_{f_0}^2(X)e^{-x^2/2}}{f_0(X)}\right]} < 1 \qquad (38)$$

since $0 < \frac{h_{f_0}^2(x)}{f_0(x)} < 1$ for all $x \in \mathcal{X}$ by construction and note

$$E\left[\frac{\left|W^K(X)'\theta_K\right|}{\sqrt{f_0(X)}}\right] \leq \sqrt{E\left[\frac{\theta_K' W^K(X)W^K(X)'\theta_K}{f_0(X)}\right]} = \sqrt{\theta_K'\theta_K} = \|\theta_K\| < 1 \qquad (39)$$

since $\|\theta_K\|^2 < 1$. From these we conclude that

$$|Q_0 - Q(\theta_K)| \leq C_1 \sup_{x\in\mathcal{X}} \frac{\sqrt{f_0(x)}}{f(x,\theta_K)} \sup_{x\in\mathcal{X}} \left|h_{f_0}(x)e^{-x^4/2} - W^K(x)'\theta_K\right| = O\left(K^{-s}\right)$$

from (10) and since $f_0(x)$ is uniformly bounded from above and $f(x,\theta_K)$ is bounded away from zero. It follows that

$$0 \leq Q(\theta_K^*) - Q(\theta_K) \leq Q(\theta_K^*) - Q_0 + C_1 K^{-s} \leq C_1 K^{-s} \qquad (40)$$

where the first inequality is by definition of $\theta_K^* = \operatorname{argmax}_\theta Q(\theta)$, the second inequality is by (36), and the last inequality is since $Q(\theta_K^*) \leq Q_0$ from the information inequality (see Gallant (1987, p.484)). Using the second order Taylor expansion where $\widetilde{\theta}$ lies between a given $\theta^o \in \Theta_n$ with $\dim(\theta^o) = K$ and θ_K^*, we have

$$\begin{aligned}
Q(\theta_K^*) - Q(\theta^o) &= -\frac{\partial Q}{\partial \theta'}(\theta_K^*)(\theta^o - \theta_K^*) - \frac{1}{2}(\theta^o - \theta_K^*)' \frac{\partial^2 Q}{\partial\theta\partial\theta'}(\widetilde{\theta})(\theta^o - \theta_K^*) \qquad (41)\\
&= -\frac{1}{2}(\theta^o - \theta_K^*)' \frac{\partial^2 Q}{\partial\theta\partial\theta'}(\widetilde{\theta})(\theta^o - \theta_K^*)
\end{aligned}$$

since $\frac{\partial Q}{\partial \theta}(\theta_K^*) = 0$ by F.O.C of (14). Consider

$$\begin{aligned}
-\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'} &= E\left[\frac{2\left(W^K(X)'\widetilde{\theta}\right)^2 W^K(X)W^K(X)'}{f(X,\widetilde{\theta})^2} - \frac{W^K(X)W^K(X)'}{f(X,\widetilde{\theta})}\right] \qquad (42)\\
&= E\left[\left(1 - 2\frac{\epsilon_0\phi(X)/\int_{\mathcal{X}}\phi(x)dx}{f(X,\widetilde{\theta})}\right)\frac{W^K(X)W^K(X)'}{f(X,\widetilde{\theta})}\right] \\
&= E\left[\left(1 - 2\frac{\epsilon_0\phi(X)/\int_{\mathcal{X}}\phi(x)dx}{f(X,\widetilde{\theta})}\right)\left(\frac{f_0(X)}{f(X,\widetilde{\theta})}\right)\frac{W^K(X)W^K(X)'}{f_0(X)}\right].
\end{aligned}$$

By construction, we can ensure $0 < 1 - 2\frac{\epsilon_0 \phi(x)/\int_{\mathcal{X}} \phi(x)dx}{f(x,\theta)} < 1$ for all $x \in \mathcal{X}$ and all $\theta \in \Theta_n$ so that $-\frac{1}{2}\frac{\partial^2 Q(\theta)}{\partial\theta\partial\theta'}$ is positive-definite for all $\theta \in \Theta_n$ by choosing ϵ_0 arbitrary small. Now note

$$\begin{aligned} \left|f_0(x) - f(x,\widetilde{\theta})\right| &\leq |f_0(x) - f(x,\theta_K)| + \left|f(x,\theta_K) - f(x,\widetilde{\theta})\right| \\ &\leq O\left(\zeta(K)K^{-s}\right) + O\left(\zeta(K)^2\right)\left(\|\theta^o - \theta_K^*\| + \|\theta_K^* - \theta_K\|\right) \end{aligned}$$

from (13) and hence (noting $f_0(x)$ is bounded away from zero and bounded from above)

$$\begin{aligned} \frac{f_0(x)}{f(x,\widetilde{\theta})} &= \frac{f_0(x)}{f(x,\widetilde{\theta}) - f_0(x) + f_0(x)} \geq \frac{f_0(x)}{\sup_{x\in\mathcal{X}}\left|f(x,\widetilde{\theta}) - f_0(x)\right| + f_0(x)} \\ &\geq \frac{f_0(x)}{O\left(\zeta(K)K^{-s}\right) + O\left(\zeta(K)^2\right)\left(\|\theta^o - \theta_K^*\| + \|\theta_K^* - \theta_K\|\right) + f_0(x)}. \end{aligned}$$

It follows that

$$\begin{aligned} -\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'} &\geq \inf_{x\in\mathcal{X},\theta\in\Theta_n}\left(1 - 2\frac{\epsilon_0\phi(x)/\int_{\mathcal{X}}\phi(x)dx}{f(x,\theta)}\right)\inf_{x\in\mathcal{X}}\frac{f_0(x)}{f(x,\widetilde{\theta})}E\left[\frac{W^K(X)W^K(X)'}{f_0(X)}\right] \\ &\geq C_1 \inf_{x\in\mathcal{X}}\left(\frac{f_0(x)}{O\left(\zeta(K)K^{-s}\right) + O\left(\zeta(K)^2\right)\left(\|\theta^o - \theta_K^*\| + \|\theta_K^* - \theta_K\|\right) + f_0(x)}\right)E\left[\frac{W^K(X)W^K(X)'}{f_0(X)}\right]. \end{aligned}$$

Finally from $E\left[\frac{W^K(X)W^K(X)'}{f_0(X)}\right] = \int_{\mathcal{X}}\frac{W^K(X)W^K(X)'}{f_0(X)}f_0(X)dx = \int_{\mathcal{X}}W^K(X)W^K(X)'dx = I_K$, we conclude

$$\lambda_{\min}\left(-\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'}\right) \geq C_1 \inf_{x\in\mathcal{X}}\left(\frac{f_0(x)}{O\left(\zeta(K)K^{-s}\right) + O\left(\zeta(K)^2\right)\left(\|\theta^o - \theta_K^*\| + \|\theta_K^* - \theta_K\|\right) + f_0(x)}\right). \quad (43)$$

From this, putting $\theta^o = \theta_K$, we note

$$\begin{aligned} \lambda_{\min}\left(-\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'}\right) &\geq C_1(1 - o(1)) \text{ if } \|\theta_K^* - \theta_K\| = o\left(\zeta(K)^{-2}\right) \\ \lambda_{\min}\left(-\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'}\right) &\geq C_2\frac{1}{O\left(\zeta(K)^2\right)\left(\|\theta_K - \theta_K^*\|\right)} \text{ otherwise.} \end{aligned}$$

Thus, from $Q(\theta_K^*) - Q(\theta_K) \geq \lambda_{\min}\left(-\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'}\right)\|\theta_K - \theta_K^*\|^2$, it follows that

$$\begin{aligned} Q(\theta_K^*) - Q(\theta_K) &\geq C_1\|\theta_K - \theta_K^*\|^2 \text{ if } \|\theta_K^* - \theta_K\| = o\left(\zeta(K)^{-2}\right) \\ Q(\theta_K^*) - Q(\theta_K) &\geq O\left(\zeta(K)^{-2}\right)\|\theta_K - \theta_K^*\| \geq O\left(\zeta(K)^{-4}\right) \text{ otherwise.} \end{aligned} \quad (44)$$

However, the case of (44) contradicts to (40) if $s > 2$, which means (40) implies $\|\theta_K^* - \theta_K\| = o\left(\zeta(K)^{-2}\right)$ under $s > 2$ and hence

$$Q(\theta_K^*) - Q(\theta_K) \geq C_1\|\theta_K - \theta_K^*\|^2. \quad (45)$$

Together with (40), it implies $C_1K^{-s} \geq Q(\theta_K^*) - Q(\theta_K) \geq C_3\|\theta_K - \theta_K^*\|^2$ and hence under $s > 2$

$$\|\theta_K - \theta_K^*\| = O\left(K^{-s/2}\right) \quad (46)$$

as claimed in Lemma 2.1. Finally note $\|\theta_K - \theta_K^*\| = o\left(\zeta(K)^{-2}\right)$ as long as (46) holds under $s > 2$. Now consider

$$
\begin{aligned}
& \sup_{x\in\mathcal{X}} |f(x,\theta_K) - f(x,\theta_K^*)| \le \sup_{x\in\mathcal{X}} \left\| \left(W^K(x)'\theta_K\right)^2 - \left(W^K(x)'\theta_K^*\right)^2 \right\| \\
\le\ & \sup_{x\in\mathcal{X}} \left\|W^K(x)'\theta_K - W^K(x)'\theta_K^*\right\| \sup_{x\in\mathcal{X}} \left\|W^K(x)'\theta_K + W^K(x)'\theta_K^*\right\| \\
\le\ & \sup_{x\in\mathcal{X}} \left\|W^K(x)\right\| \|\theta_K - \theta_K^*\| \sup_{x\in\mathcal{X}} \left\|W^K(x)\right\| \left(\|\theta_K\| + \|\theta_K^*\|\right) = O\left(\zeta(K)^2 K^{-s/2}\right)
\end{aligned}
$$

from the Cauchy-Schwarz inequality, (46), $\sup_{x\in\mathcal{X}} \left\|W^K(x)\right\| \le \zeta(K)$, and $\|\theta\|^2 < 1$ for any $\theta \in \Theta_n$. It follows that

$$
\begin{aligned}
& \sup_{x\in\mathcal{X}} |f_0(x) - f_K^*(x)| \le \sup_{x\in\mathcal{X}} |f_0(x) - f(x,\theta_K)| + \sup_{x\in\mathcal{X}} |f(x,\theta_K) - f(x,\theta_K^*)| \\
\le\ & O\left(\zeta(K)K^{-s}\right) + O\left(\zeta(K)^2 K^{-s/2}\right) = O\left(\zeta(K)^2 K^{-s/2}\right).
\end{aligned}
$$

D Uniform Law of Large Numbers

Now Define $\widehat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \ln f(X_i,\theta)$ and $Q(\theta) = E\left[\ln f(X,\theta)\right]$ where $\theta \in \Theta_n$ defined in (4). We establish a uniform convergence with rate as

$$
\sup_{\theta\in\Theta_n} \left|\widehat{Q}_n(\theta) - Q(\theta)\right| = o_p\left(n^{-1/2+\alpha/2+\delta}\right)
$$

following Lemma 2 in Fenton and Gallant (1996a), which is a variant of results obtained in White and Wooldrige (1991), Gallant and Souza (1991), De Jong (1993) among others.

Lemma D.1 (Lemma 2 in Fenton and Gallant (1996a)) *Let $\{\Theta_n\}$ be a sequence of compact subsets of a metric space (Θ,ρ). Let $\{s_{ni}(\theta) : \theta \in \Theta;\ i = 1,\ldots,n;\ n = 1,\ldots\}$ be a sequence of real valued random variables defined over a complete probability space $(\Omega,\mathcal{A},P)$. Suppose that there are sequences of positive numbers $\{d_n\}$ and $\{M_n\}$ such that for each θ^o in Θ_n and for all θ in $\eta_n(\theta^o) = \{\theta \in \Theta_n : \rho(\theta,\theta^o) < d_n\}$, we have $|s_{ni}(\theta) - s_{ni}(\theta^o)| \le \frac{1}{n}M_n\rho(\theta,\theta^o)$. Let $G_n(\tau)$ be the smallest number of open balls of radius τ necessary to cover Θ_n. If $\sup_{\theta\in\Theta_n} \Pr\left\{\left|\sum_{i=1}^n \left(s_{ni}(\theta) - E\left[s_{ni}(\theta)\right]\right)\right| > \epsilon\right\} \le \Gamma_n(\epsilon)$, then for all sufficiently small $\varepsilon > 0$ and all sufficiently large n,*

$$
\Pr\left\{\sup_{\theta\in\Theta_n} \left|\sum_{i=1}^n \left(s_{ni}(\theta) - E\left[s_{ni}(\theta)\right]\right)\right| > \varepsilon M_n d_n\right\} \le G_n\left(\frac{\varepsilon d_n}{3}\right)\Gamma_n\left(\frac{\varepsilon M_n d_n}{3}\right).
$$

Now define $s_{ni}(\theta) = \frac{1}{n}\ln f\left(X_i,\theta\right)$. Then, we have $\widehat{Q}_n(\theta) = \sum_{i=1}^n s_{ni}(\theta)$ and $Q(\theta) = \sum_{i=1}^n E\left[s_{ni}(\theta)\right]$. To entertain Lemma (D.1), in what follows, three conditions of Lemma (D.1) are verified.

Lemma D.2 *Suppose Assumption 2.2 holds. Then, $|s_{ni}(\theta) - s_{ni}(\theta^o)| \le C\frac{1}{n}\zeta(K(n))^2 \|\theta - \theta^o\|$.*

Proof. Note if $0 < c \le a \le b$, then $|\ln a - \ln b| \le |a - b|/c$. Since $f(x,\theta)$ is bounded away from zero for all $\theta \in \Theta_n$ and $x \in \mathcal{X}$, we have $0 < C \le f(x,\theta)$ from which it follows that

$$
|s_{ni}(\theta) - s_{ni}(\theta^o)| \le |f(X_i,\theta) - f(X_i,\theta^o)|/nC.
$$

Consider

$$\begin{aligned} & |f(X_i,\theta)-f(X_i,\theta^o)| = \left|W^K(X_i)'\theta + W^K(X_i)'\theta^o\right| \left|W^K(X_i)'\theta - W^K(X_i)'\theta^o\right| \\ \leq & \sup_{x\in\mathcal{X}} \left\|W^K(x)\right\| (\|\theta\| + \|\theta^o\|) \sup_{x\in\mathcal{X}} \left\|W^K(x)\right\| \|\theta-\theta^o\| \leq C_1\zeta(K)^2 \|\theta-\theta^o\| \end{aligned}$$

since $\|\theta\| < 1$ for all $\theta \in \Theta_n$ and $\sup_{x\in\mathcal{X}} \left\|W^K(x)\right\| = \zeta(K)$. It follows that

$$|s_{ni}(\theta) - s_{ni}(\theta^o)| \leq |f(X_i,\theta) - f(X_i,\theta^o)| /nC \leq C_2\zeta(K)^2 \|\theta-\theta^o\| /n.$$

■

Lemma D.3 *Suppose Assumption 2.2 holds and* $\zeta(K) = O\left(K^\zeta\right)$. *Then,*

$$\Pr\left\{\left|\sum_{i=1}^n (s_{ni}(\theta) - E[s_{ni}(\theta)])\right| > \epsilon\right\} \leq 2\exp\left(\frac{-2\epsilon^2}{n\left(\frac{1}{n}2\zeta\ln K(n) + \frac{1}{n}C\right)^2}\right).$$

Proof. We have $0 < C_1 \leq f(x,\theta) < C_2K^{2\zeta} + \epsilon_0\frac{\phi(0)}{\int_\mathcal{X}\phi(x)dx}$ by construction and since $f(x,\theta)$ is bounded away from zero. Thus it follows that $\frac{1}{n}C_3 < s_{ni}(\theta) < \frac{1}{n}2\zeta\ln K + \frac{1}{n}C_4$ for sufficiently large K. Hoeffding's (1963) inequality implies that $\Pr\left(|Y_1 + \ldots + Y_n| \geq \epsilon\right) \leq 2\exp\left(-2\epsilon^2/\sum_{i=1}^n (b_i - a_i)^2\right)$ for independent random variables centered zero with ranges $a_i \leq Y_i \leq b_i$. Applying this inequality, we have

$$\Pr\left\{\left|\sum_{i=1}^n (s_{ni}(\theta) - E[s_{ni}(\theta)])\right| > \epsilon\right\} \leq 2\exp\left(-2\epsilon^2/n\left(\frac{1}{n}2\zeta\ln K(n) + \frac{1}{n}C\right)^2\right)$$

■

Lemma D.4 (Lemma 6 in Fenton and Gallant(1996a)) *The number of open balls of radius* δ *required to cover* Θ_n *is bounded by* $2K(n)(2/\delta+1)^{K(n)-1}$.
Proof. *Lemma 1 of Gallant and Souza (1991) shows that the number of radius-*δ *balls needed to cover the surface of a unit sphere in* $\mathcal{R}^p$ *is bounded by* $2p(2/\delta+1)^{p-1}$. *Noting* $\dim(\Theta_n) = K(n)$, *the result follows immediately.* ■

Applying the results of Lemma D.2-D.4, finally we obtain

Lemma D.5 *Let* $K(n) = Cn^\alpha$ *with* $\alpha \in (0,1)$ *and suppose Assumption 2.2 holds.*
Then, $\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - Q(\theta)\right| = o_p\left(n^{-1/2+\alpha/2+\delta}\right)$.

Proof. Let $M_n = C_1O\left(K^{2\zeta}\right) = C_2n^{2\zeta\alpha}$, $d_n = \frac{1}{C_1}n^{-(2\zeta-1)\alpha-\beta}$, and $\rho(\theta,\theta^o) = \|\theta-\theta^o\|$. Then from Lemma D.1, we have

$$\begin{aligned} & \Pr\left\{\sup_{\theta\in\Theta_n}\left|\sum_{i=1}^n (s_{ni}(\theta) - E[s_{ni}(\theta)])\right| > \varepsilon n^{\alpha-\beta}\right\} \\ \leq & 4C\cdot n^\alpha(\frac{6C_1}{\varepsilon}n^{(2\zeta-1)\alpha+\beta} + 1)^{n^\alpha-1}\exp\left(-2\left(\frac{\varepsilon n^{\alpha-\beta}}{3}\right)^2/n\left(\frac{1}{n}2\zeta\ln K(n) + \frac{1}{n}C_2\right)^2\right) \end{aligned}$$

applying Lemma D.2, Lemma D.3, and Lemma D.4.

Note $4C \cdot n^{\alpha}(\frac{6C_1}{\varepsilon} n^{(2\zeta-1)\alpha+\beta} + 1)^{n^{\alpha}-1}$ is dominated by $C_3 n^{\alpha} n^{((2\zeta-1)\alpha+\beta)(n^{\alpha}-1)}$ for sufficiently large n and note $n\left(\frac{1}{n}2\zeta \ln K(n) + \frac{1}{n}C_2\right)^2$ is dominated by $n\left(\frac{1}{n}2\zeta \ln K(n)\right)^2$. Thus, we simplify

$$\begin{aligned} & \Pr\left\{\sup_{\theta\in\Theta_n}\left|\sum_{i=1}^{n}\left(s_{ni}(\theta) - E\left[s_{ni}(\theta)\right]\right)\right| > \varepsilon n^{\alpha-\beta}\right\} \\ \leq\ & C_4 \exp\left(\ln\left(n^{\alpha} n^{((2\zeta-1)\alpha+\beta)(n^{\alpha}-1)}\right) - \frac{2\varepsilon^2}{9} n^{2\alpha-2\beta+1} / \left(2\zeta \ln K(n)\right)^2\right) \\ =\ & C_4 \exp\left(\alpha \ln n + \left((2\zeta-1)\alpha+\beta\right)\left(n^{\alpha}-1\right)\ln n - \frac{2\varepsilon^2}{9} n^{2\alpha-2\beta+1} / \left(2\zeta \ln K(n)\right)^2\right) \end{aligned}$$

for sufficiently large n. As long as $2\alpha - 2\beta + 1 > \alpha$, $\frac{2\varepsilon^2}{9} n^{2\alpha-2\beta+1} / \left(2\zeta \ln K(n)\right)^2$ dominates $\alpha \ln n + \left((2\zeta-1)\alpha+\beta\right)\left(n^{\alpha}-1\right)\ln n$ and hence we conclude

$$\Pr\left\{\sup_{\theta\in\Theta_n}\left|\sum_{i=1}^{n}\left(s_{ni}(\theta) - E\left[s_{ni}(\theta)\right]\right)\right| > \varepsilon n^{\alpha-\beta}\right\} = o(1)$$

provided that $\frac{\alpha+1}{2} > \beta > \alpha$. By taking $\beta = \frac{1}{2} + \frac{1}{2}\alpha - \delta$ (the best possible rate), we have

$$\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - Q(\theta)\right| = \sup_{\theta\in\Theta_n}\left|\sum_{i=1}^{n}\left(s_{ni}(\theta) - E\left[s_{ni}(\theta)\right]\right)\right| = o_p(n^{-\frac{1}{2}+\frac{1}{2}\alpha+\delta})$$

for all sufficiently small $\delta > 0$. ■

Lemma D.6 *Suppose (i) Assumption 2.2 holds and (ii)* $\frac{\zeta(K)^2 K}{n} \to 0$. *Let* $\eta_n = n^{-\beta_\eta}$ *with* $\beta_\eta \leq 1/2 - \alpha/2 - \delta$, *then*

$$\sup_{\|\theta-\theta^o\|\leq o(\eta_n),\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta^o) - \left(Q(\theta) - Q(\theta^o)\right)\right| = o_p\left(\eta_n n^{-1/2+\alpha/2+\delta}\right)$$

Proof. Applying the mean value theorem for $\widetilde{\theta}$ that lies between θ and θ^o, we can rewrite

$$\begin{aligned} & \widehat{Q}_n(\theta) - \widehat{Q}_n(\theta^o) - \left(Q(\theta) - Q(\theta^o)\right) \\ =\ & \widehat{Q}_n(\theta) - Q(\theta) - \left(\widehat{Q}_n(\theta^o) - Q(\theta^o)\right) = \left(\frac{\partial\left(\widehat{Q}_n(\widetilde{\theta}) - Q(\widetilde{\theta})\right)}{\partial\theta'}\right)\left(\theta - \theta^o\right). \end{aligned} \tag{47}$$

Now consider for any $\overline{\theta} \in \Theta_n$,

$$\frac{\partial\widehat{Q}_n(\overline{\theta})}{\partial\theta} - \frac{\partial Q(\overline{\theta})}{\partial\theta} = 2\frac{1}{n}\sum_{i=1}^{n}\left(\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)} - E\left[\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right]\right) \tag{48}$$

and define $L_{ni} = \left(\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)} - E\left[\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right]\right)$. Noting L_{ni} is a triangular array of i.i.d random variables with mean zero, we bound (48) as follows. First consider

$$\begin{aligned} \mathrm{Var}\left[L_{ni}\right] \leq\ & E\left[L_{ni}L_{ni}'\right] = E\left[\left(\frac{W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right)^2 W^K(X_i)W^K(X_i)'\right] \\ & -E\left[\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right] E\left[\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right]'. \end{aligned} \tag{49}$$

The first term in the right hand side of (49) is bounded by

$$\begin{aligned}
& E\left[\left(\frac{W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right)^2 W^K(X_i)W^K(X_i)'\right] \\
= \; & E\left[\left(1-\frac{\epsilon_0\phi(X_i)/\int_{\mathcal{X}}\phi(x)dx}{f\left(X_i,\overline{\theta}\right)}\right)\left(\frac{f_0(X_i)}{f\left(X_i,\overline{\theta}\right)}\right)\frac{W^K(X_i)W^K(X_i)'}{f_0(X_i)}\right] \\
\leq \; & \sup_{x\in\mathcal{X},\overline{\theta}\in\Theta_n}\left(\left(1-\frac{\epsilon_0\phi(x)/\int_{\mathcal{X}}\phi(x)dx}{f\left(x,\overline{\theta}\right)}\right)\frac{f_0(x)}{f\left(x,\overline{\theta}\right)}\right)E\left[\frac{W^K(X_i)W^K(X_i)'}{f_0(X_i)}\right] \\
\leq \; & \sup_{x\in\mathcal{X},\overline{\theta}\in\Theta_n}\left(\frac{f_0(x)}{f\left(x,\overline{\theta}\right)}\right)E\left[\frac{W^K(X_i)W^K(X_i)'}{f_0(X_i)}\right] = C_1 I_K
\end{aligned}$$

since $f\left(x,\overline{\theta}\right)$ is bounded away from zero uniformly over $\mathcal{X}$ and $\overline{\theta}\in\Theta_n$ and since $f_0(x)$ is bounded away from above. Next consider, for the second term

$$\begin{aligned}
& E\left[\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right]E\left[\frac{W^K(X_i)W^K(X_i)'\overline{\theta}}{f\left(X_i,\overline{\theta}\right)}\right]' = E\left[\frac{W^K(X_i)W^K(X_i)'}{f\left(X_i,\overline{\theta}\right)}\right]\overline{\theta}\overline{\theta}'E\left[\frac{W^K(X_i)W^K(X_i)'}{f\left(X_i,\overline{\theta}\right)}\right] \\
\leq \; & \left(\sup_{x\in\mathcal{X},\overline{\theta}\in\Theta_n}\left(\frac{f_0(x)}{f\left(x,\overline{\theta}\right)}\right)\right)^2 E\left[\frac{W^K(X_i)W^K(X_i)'}{f_0(X_i)}\right]\overline{\theta}\overline{\theta}'E\left[\frac{W^K(X_i)W^K(X_i)'}{f_0(X_i)}\right] \leq C_2\overline{\theta}\overline{\theta}'
\end{aligned}$$

since $\sup_{x\in\mathcal{X},\overline{\theta}\in\Theta_n}\left(\frac{f_0(x)}{f\left(x,\overline{\theta}\right)}\right)<\infty$ as before. Now note

$$\begin{aligned}
E\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n L_{ni}\right\|\right) & \leq \sqrt{\mathrm{tr}\left(\mathrm{Var}\left[L_{ni}\right]\right)} \leq \sqrt{C_1\mathrm{tr}\left(I_K\right)+C_2\mathrm{tr}\left(\overline{\theta}\overline{\theta}'\right)} \\
& = \sqrt{C_1 K + C_2\left\|\overline{\theta}\right\|^2} = O(\sqrt{K})
\end{aligned}$$

uniformly over $\overline{\theta}\in\Theta_n$ since $\left\|\overline{\theta}\right\|^2<1$ for any $\overline{\theta}\in\Theta_n$ by construction. It follows that

$$\sup_{\overline{\theta}\in\Theta_n}\left\|\frac{\partial\widehat{Q}_n(\overline{\theta})}{\partial\theta}-\frac{\partial Q(\overline{\theta})}{\partial\theta}\right\| = 2\frac{1}{\sqrt{n}}\sup_{\overline{\theta}\in\Theta_n}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n L_{ni}\right\| = O_p\left(\sqrt{\frac{K}{n}}\right)$$

from the Markov inequality. Thus, noting $O_p\left(\sqrt{\frac{K}{n}}\right) = o_p\left(n^{-1/2+\alpha/2+\delta}\right)$ for $K=n^\alpha$ and sufficiently small δ, we have

$$\begin{aligned}
& \sup_{\|\theta-\theta^o\|\leq o(\eta_n),\theta\in\Theta_n}\left\|\widehat{Q}_n(\theta)-\widehat{Q}_n(\theta^o)-(Q(\theta)-Q(\theta^o))\right\| \\
\leq \; & \sup_{\theta\in\Theta_n}\left\|\frac{\partial\left(\widehat{Q}_n(\theta)-Q(\theta)\right)}{\partial\theta}\right\|\sup_{\|\theta-\theta^o\|\leq o(\eta_n),\theta\in\Theta_n}\|\theta-\theta^o\| = o_p\left(\eta_n n^{-1/2+\alpha/2+\delta}\right)
\end{aligned}$$

from (47) applying the Cauchy-Schwarz inequality. ■

E Proof of Lemma 2.2

First, from (43) and $\|\theta_K^* - \theta_K\| = O\left(K^{-s/2}\right)$ with $s > 2$, we note that

$$\lambda_{\min}\left(-\frac{1}{2}\frac{\partial^2 Q(\widetilde{\theta})}{\partial\theta\partial\theta'}\right) \geq C_1 \inf_{x\in\mathcal{X}}\left(\frac{f_0(x)}{O\left(\zeta(K)K^{-s}\right) + O\left(\zeta(K)^2\right)\left(\|\theta-\theta_K^*\|\right) + f_0(x)}\right)$$

where $\widetilde{\theta}$ lies between θ and θ_K^* and hence from (41),

$$\begin{aligned} Q(\theta_K^*) - Q(\theta) &\geq C_1\|\theta-\theta_K^*\|^2 \text{ if } \|\theta-\theta_K^*\| = o\left(\zeta(K)^{-2}\right) \\ Q(\theta_K^*) - Q(\theta) &\geq C_2\zeta(K)^{-2}\|\theta-\theta_K^*\| \text{ otherwise.} \end{aligned} \tag{50}$$

Denote $\kappa = 1/2 - \alpha/2 - \delta$ and $\eta_{0n} = o(n^{-\kappa})$. We derive the convergence rate in two cases: one is when η_{0n} has the equal or a smaller order than $o\left(\zeta(K)^{-4}\right)$ and the other case is when η_{0n} has a larger order than $o\left(\zeta(K)^{-4}\right)$.

1) *When η_{0n} has equal or smaller order than $o\left(\zeta(K)^{-4}\right)$*, which holds under $\alpha < \frac{1}{5}$:

Now let $\delta_{0n} = \sqrt{2\eta_{0n}}$. For any c such that $C_1c^2 > 1$, it follows

$$\Pr\left(\left\|\widehat{\theta}_K - \theta_K^*\right\| \geq c\delta_{0n}\right) \leq \Pr\left(\sup_{\left\|\theta-\theta_K^*\right\|\geq c\delta_{0n},\theta\in\Theta_n}\widehat{Q}_n(\theta) \geq \widehat{Q}_n(\theta_K^*)\right) \tag{51}$$

$$\begin{aligned} &\leq \Pr\left(\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - Q(\theta)\right| > \eta_{0n}\right) \\ &\quad + \Pr\left(\left\{\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - Q(\theta)\right| \leq \eta_{0n}\right\} \cap \left\{\sup_{\left\|\theta-\theta_K^*\right\|\geq c\delta_{0n},\theta\in\Theta_n}\widehat{Q}_n(\theta) \geq \widehat{Q}_n(\theta_K^*)\right\}\right) \\ &\leq \Pr\left(\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - Q(\theta)\right| > \eta_{0n}\right) + \Pr\left(\sup_{\left\|\theta-\theta_K^*\right\|\geq c\delta_{0n},\theta\in\Theta_n} Q(\theta) \geq Q(\theta_K^*) - 2\eta_{0n}\right) = P_1 + P_2. \end{aligned}$$

Now note $P_1 \to 0$ from (16). Now we show $P_2 \to 0$. This holds since $Q(\theta)$ has its maximum at θ_K^*. To be precise, note

$$\begin{aligned} Q(\theta_K^*) - Q(\theta) &\geq C_1\|\theta-\theta_K^*\|^2 \geq 2C_1c^2\eta_{0n} \text{ if } \|\theta-\theta_K^*\| = o\left(\zeta(K)^{-2}\right) \\ Q(\theta_K^*) - Q(\theta) &\geq C_2\zeta(K)^{-2}\|\theta-\theta_K^*\| \geq C_3\zeta(K)^{-4} \text{ otherwise} \end{aligned}$$

and hence

$$\begin{aligned} \sup_{\left\|\theta-\theta_K^*\right\|\geq c\delta_{0n},\theta\in\Theta_n} Q(\theta) - Q(\theta_K^*) &\leq -2C_1c^2\eta_{0n} \text{ if } \|\theta-\theta_K^*\| = o\left(\zeta(K)^{-2}\right) \\ \sup_{\left\|\theta-\theta_K^*\right\|\geq c\delta_{0n},\theta\in\Theta_n} Q(\theta) - Q(\theta_K^*) &< -C_3\zeta(K)^{-4} \text{ otherwise.} \end{aligned}$$

Therefore, as long as $C_1c^2 > 1$ and $\zeta(K)^4\eta_{0n} \to 0$, we have $P_2 \to 0$. $\zeta(K)^4\eta_{0n} \to 0$ holds under $\alpha < \frac{1}{5}$. Thus, we have proved $\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-\kappa/2})$. Now we refine the convergence rate by exploiting the local curvature of $\widehat{Q}_n(\theta)$ around θ_K^*. Let $\eta_{1n} = n^{-\kappa}\delta_{0n} = o\left(n^{-(\kappa+\kappa/2)}\right)$ and $\delta_{1n} = \sqrt{\eta_{1n}} = o\left(n^{-(\kappa/2+\kappa/4)}\right)$. For any c such that $C_1c^2 > 1$, we have

$$\Pr\left(\left\|\widehat{\theta}_K - \theta_K^*\right\| \geq c\delta_{1n}\right) \leq \Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \widehat{Q}_n(\theta) \geq \widehat{Q}_n(\theta_K^*)\right) \tag{52}$$

$$\begin{aligned}
\leq\ & \Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) - (Q(\theta) - Q(\theta_K^*))\right| > \eta_{1n}\right) \\
& + \Pr\left(\begin{array}{c} \left\{\sup\limits_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) - (Q(\theta) - Q(\theta_K^*))\right| \leq \eta_{1n}\right\} \\ \cap\left\{\sup\limits_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \widehat{Q}_n(\theta) \geq \widehat{Q}_n(\theta_K^*)\right\}\end{array}\right) \\
\leq\ & \Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) - (Q(\theta) - Q(\theta_K^*))\right| > \eta_{1n}\right) \\
& + \Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} Q(\theta) \geq Q(\theta_K^*) - \eta_{1n}\right) = P_3 + P_4
\end{aligned}$$

where[9] $P_3 \to 0$ from Lemma D.6 and $P_4 \to 0$ similarly with P_2 noting

$$\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} Q(\theta) - Q(\theta_K^*) \leq -C_1c^2\eta_{1n}$$

by (50) and since $\|\theta - \theta_K^*\| = o\left(\zeta(K)^{-2}\right)$ for any θ such that $\delta_{0n} \geq \|\theta - \theta_K^*\| \geq c\delta_{1n}$ under $\alpha < \frac{1}{5}$. This show that $\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-(\kappa/2+\kappa/4)})$. Repeating this refinement for infinite number of times, we obtain

$$\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-(\kappa/2+\kappa/4+\kappa/8+\ldots)}) = o_p(n^{-\kappa}) = o_p(n^{-1/2+\alpha/2+\delta})$$

---

[9] Note

$$\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) - (Q(\theta) - Q(\theta_K^*))\right| \leq \eta_{1n}$$

implies, for any θ such that $\delta_{0n} \geq \left\|\theta - \theta_K^*\right\| \geq c\delta_{1n}$,

$$\begin{aligned}
-\eta_{1n} - \sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} (Q(\theta) - Q(\theta_K^*)) \quad &\leq \quad \widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) \\
&\leq \quad \eta_{1n} + \sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} (Q(\theta) - Q(\theta_K^*))
\end{aligned}$$

and hence we obtain

$$\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) \leq \eta_{1n} + \sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} (Q(\theta) - Q(\theta_K^*))$$

Therefore,

$$\Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} \widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) > 0\right) \leq \Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} Q(\theta) \geq Q(\theta_K^*) - \eta_{1n}\right),$$

from which we have obtained the third inequality.

under $\alpha < \frac{1}{5}$.

2) Now we consider *when* η_{0n} *has larger order than* $o\left(\zeta(K)^{-4}\right)$ (which holds under $\alpha \geq \frac{1}{5}$):

Let $\widetilde{\delta}_{0n} = o\left(\zeta(K)^{-2}\right) n^{\beta}$ for $\beta > 0$. Then, from (51), we have

$$\begin{aligned} & \Pr\left(\left\|\widehat{\theta}_K - \theta_K^*\right\| \geq c\widetilde{\delta}_{0n}\right) \\ \leq \; & \Pr\left(\sup_{\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - Q(\theta)\right| > \eta_{0n}\right) + \Pr\left(\sup_{\left\|\theta-\theta_K^*\right\|\geq c\widetilde{\delta}_{0n},\theta\in\Theta_n} Q(\theta) \geq Q(\theta_K^*) - 2\eta_{0n}\right) = P_1 + P_2. \end{aligned}$$

Again note $P_1 \to 0$ from (16). Now we show $P_2 \to 0$. Note from (50),

$$\sup_{\left\|\theta-\theta_K^*\right\|\geq c\widetilde{\delta}_{0n},\theta\in\Theta_n} Q(\theta) - Q(\theta_K^*) \leq -C_2\zeta(K)^{-2}\left\|\theta - \theta_K^*\right\| \leq -o\left(\zeta(K)^{-4}\right) n^{\beta}$$

since $\|\theta - \theta_K^*\| > o\left(\zeta(K)^{-2}\right)$ for any θ such that $\|\theta - \theta_K^*\| \geq c\widetilde{\delta}_{0n}$. It follows that $P_2 \to 0$ as long as $\zeta(K)^4 n^{-\beta}\eta_{0n} \to 0$, which holds under

$$\beta > \frac{5}{2}\alpha - \frac{1}{2} + \delta \tag{53}$$

and hence the convergence rate will be $o_p\left(\delta_{0n}\right) = o_p\left(n^{-\alpha+\beta}\right)$. Now we refine the convergence rate by exploiting the local curvature of $\widehat{Q}_n(\theta)$ around θ_K^* again. Let $\widetilde{\eta}_{1n} = n^{-\kappa}\widetilde{\delta}_{0n} = o\left(n^{-((\alpha-\beta)+\kappa)}\right)$ and $\widetilde{\delta}_{1n} = \sqrt{\widetilde{\eta}_{1n}} = o\left(n^{-((\alpha-\beta)/2+\kappa/2)}\right)$. Then, from (52), we have

$$\begin{aligned} & \Pr\left(\left\|\widehat{\theta}_K - \theta_K^*\right\| \geq c\widetilde{\delta}_{1n}\right) \\ \leq \; & \Pr\left(\sup_{\widetilde{\delta}_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\widetilde{\delta}_{1n},\theta\in\Theta_n}\left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) - (Q(\theta) - Q(\theta_K^*))\right| > \widetilde{\eta}_{1n}\right) \\ & + \Pr\left(\sup_{\delta_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\delta_{1n},\theta\in\Theta_n} Q(\theta) \geq Q(\theta_K^*) - \widetilde{\eta}_{1n}\right) = P_3 + P_{41} \end{aligned} \tag{54}$$

where $P_3 \to 0$ from Lemma D.6. Now we show $P_{41} \to 0$ similarly with P_2. Here again we need to consider two cases:

2-1) *When* $\widetilde{\delta}_{1n}$ *has equal or smaller order than* $o\left(\zeta(K)^{-2}\right)$, which holds under $\beta \leq \frac{1}{2} - \frac{3}{2}\alpha - \delta$ and hence from $\alpha > \beta$ and (53) it requires $1/5 \leq \alpha < 1/4$. Under this case, note

$$\begin{aligned} \sup_{\widetilde{\delta}_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\widetilde{\delta}_{1n},\theta\in\Theta_n} Q(\theta) - Q(\theta_K^*) & \leq -C_1\left\|\theta - \theta_K^*\right\|^2 \leq -C_1c^2\widetilde{\delta}_{1n}^2 = -C_1c^2\widetilde{\eta}_{1n} \text{ if } \left\|\theta - \theta_K^*\right\| = o\left(\zeta(K)^{-2}\right) \\ \sup_{\widetilde{\delta}_{0n}\geq\left\|\theta-\theta_K^*\right\|\geq c\widetilde{\delta}_{1n},\theta\in\Theta_n} Q(\theta) - Q(\theta_K^*) & \leq -C_2\zeta(K)^{-2}\left\|\theta - \theta_K^*\right\| \leq -C_2\zeta(K)^{-4} \text{ otherwise} \end{aligned}$$

by (50) and hence $P_{41} \to 0$ as long as $C_1c^2 > 1$ and $\zeta(K)^4\widetilde{\eta}_{1n} = \zeta(K)^4\widetilde{\eta}_{1n}n^{-\kappa}\widetilde{\delta}_{0n} = \zeta(K)^2 o\left(n^{-\kappa}n^{\beta}\right) \to 0$, which holds under $\beta \leq \frac{1}{2} - \frac{3}{2}\alpha - \delta$. Repeating this refinement for infinite number of times (noting that for any θ such that $\widetilde{\delta}_{1n} \geq \|\theta - \theta_K^*\|$, we have $\|\theta - \theta_K^*\| = o\left(\zeta(K)^{-2}\right)$), we obtain

$$\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-\left(\lim_{L\to\infty}\left(\sum_{l=1}^{L}\frac{\kappa}{2^l} + \frac{(\alpha-\beta)}{L}\right)\right)}) = o_p(n^{-\kappa})$$

and hence the effect of η_{0n}'s *having larger order than* $o\left(\zeta(K)^{-4}\right)$ disappear ($\frac{(\alpha-\beta)}{L}$ goes to zero as L goes to infinity)). This makes sense because the iterated convergence rate improvement using the local curvature will dominate the convergence rate from the uniform convergence.

2-2) *When* $\widetilde{\delta}_{1n}$ *has bigger order than* $o\left(\zeta(K)^{-2}\right)$, which holds under $\beta > \frac{1}{2} - \frac{3}{2}\alpha - \delta$ and $1/3 > \alpha \geq 1/4$: In this case, we let $\widetilde{\delta}_{1n} = \widetilde{\delta}_{0n} n^{-\gamma}$ for some $\gamma > 0$ and hence we require $\beta > \gamma$. From (54), we note

$$
\begin{aligned}
& \Pr\left(\left\|\widehat{\theta}_K - \theta_K^*\right\| \geq c\widetilde{\delta}_{1n}\right) \\
\leq\ & \Pr\left(\sup_{\widetilde{\delta}_{0n} \geq \left\|\theta-\theta_K^*\right\| \geq c\widetilde{\delta}_{1n}, \theta\in\Theta_n} \left|\widehat{Q}_n(\theta) - \widehat{Q}_n(\theta_K^*) - (Q(\theta) - Q(\theta_K^*))\right| > \widetilde{\eta}_{1n}\right) \\
& + \Pr\left(\sup_{\widetilde{\delta}_{0n} \geq \left\|\theta-\theta_K^*\right\| \geq c\widetilde{\delta}_{1n}, \theta\in\Theta_n} Q(\theta) \geq Q(\theta_K^*) - \widetilde{\eta}_{1n}\right) = P_3 + P_{42}.
\end{aligned}
$$

We have seen that P_3 goes to zero since $\widetilde{\eta}_{1n} = n^{-\kappa}\widetilde{\delta}_{0n}$ and by Lemma D.6. Now we verify P_{42} goes to zero. From (50), to have $P_{42} \to 0$, we require that $\zeta(K)^{-2}\widetilde{\delta}_{1n}$ have a bigger order than $\widetilde{\eta}_{1n}$ and hence we need $\gamma < \frac{1}{2} - \frac{3\alpha}{2} - \delta$. Now we improve the convergence rate again using the local curvature by defining $\widetilde{\eta}_{2n} = n^{-\kappa}\widetilde{\delta}_{1n} = o\left(n^{-((\alpha-\beta+\gamma)+\kappa)}\right)$ and $\widetilde{\delta}_{2n} = \sqrt{\widetilde{\eta}_{2n}} = o\left(n^{-((\alpha-\beta+\gamma)/2+\kappa/2)}\right)$. Then, similarly with before, at the end, we will obtain $\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-\kappa})$ as long as $\widetilde{\delta}_{2n}$ has equal or smaller order than $o\left(\zeta(K)^{-2}\right)$. The tricky case is again when $\widetilde{\delta}_{2n}$ has a bigger order than $o\left(\zeta(K)^{-2}\right)$, which happens when $\beta - \gamma > \frac{1}{2} - \frac{3}{2}\alpha - \delta$ but applying the same trick, at the end, we will obtain the same convergence rate of $\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-\kappa})$ as long as $1/3 > \alpha$. Combining these results, we conclude that under $\alpha < 1/3$, we have

$$\left\|\widehat{\theta}_K - \theta_K^*\right\| = o_p(n^{-\kappa}) = o_p(n^{-1/2+\alpha/2+\delta}).$$

This result is intuitive in the sense that ignoring δ, we obtain $o\left(\zeta\left(K\right)^{-2}\right) = o(n^{-\kappa}) = n^{-1/3}$ at $\alpha = 1/3$ and hence if $\alpha \geq 1/3$, there is no room to improve the convergence rate using the local curvature.

F Proof of Lemma 3.1

Note

$$
\begin{aligned}
& \frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln f(x_i) \\
=\ & \frac{1}{n_\gamma - 1}\sum_{i \text{ odd}}(1+\gamma)\ln f(x_i) + \frac{1}{n_\gamma - 1}\sum_{i \text{ even}}(1-\gamma)\ln f(x_i) + O_p\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i \notin \mathcal{N}_x)\right) \\
=\ & \frac{1}{2}(1+\gamma)E_f\left[\ln f(x_i)\right] + o_p(1) + \frac{1}{2}(1-\gamma)E_f\left[\ln f(x_i)\right] + o_p(1) + o_p(1) = E_f\left[\ln f(x_i)\right] + o_p(1)
\end{aligned}
$$

by the law of large numbers under Condition (i) and by Condition (iii). Similarly we have

$$
\begin{aligned}
\frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_x} c_i(\gamma)\ln g(x_{i+1}) &= E_f\left[\ln g(x_{i+1})\right] + o_p(1),\ \frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln g(y_i) = E_g\left[\ln g(y_i)\right] + o_p(1) \\
\frac{1}{n_\gamma - 1}\sum_{i\in\mathcal{N}_y} c_i(\gamma)\ln f(y_{i+1}) &= E_g\left[\ln f(y_{i+1})\right] + o_p(1)
\end{aligned}
$$

and thus $\widetilde{I}_\gamma \underset{p}{\longrightarrow} I$ noting $I(f,g)$ can be expressed as $I(f,g) = E_f[\ln f(x_i) - \ln g(x_{i+1})] + E_g[\ln g(y_i) - \ln f(y_{i+1})]$. Condition (iv) implies $\widehat{I}_\gamma - \widetilde{I}_\gamma \underset{p}{\longrightarrow} 0$. From these, we conclude $\widehat{I}_\gamma \underset{p}{\longrightarrow} I$.

G Convergence Rate of the Log Density

Lemma G.1 *Suppose Assumption 2.2 holds, $\frac{\zeta(K)^2 K}{n} \to 0$, and $\widehat{f}$ is the SNP density estimator from the sample $\{x_i\}_{i=1}^n$. Then,*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\ln \widehat{f}(x_i) - \ln f_0(x_i)\right) = o_p\left(n^{-1+\alpha+2\delta}\right) + O(K^{-s}).$$

Now consider an additional sample $\{v_i\}_{i=1}^{n_v}$ where the true density of v_i is also f_0. $\{v_i\}_{i=1}^{n_v}$ can be a subsample of $\{x_i\}_{i=1}^n$ or partly contains some of $\{x_i\}_{i=1}^n$ or totally independent with $\{x_i\}_{i=1}^n$. Provided that $n_v = \rho n$ for some positive constant ρ, we have

$$\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\ln \widehat{f}(v_i) - \ln f_0(v_i)\right) = O\left(\zeta(K)^2\right) o_p\left(n^{-1+\alpha+2\delta}\right) + O(K^{-s})$$

Thus, under $\alpha \le \frac{1}{4} - \delta$ and $s \ge \frac{1}{2\alpha}$,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\ln \widehat{f}(x_i) - \ln f_0(x_i)\right) = o_p\left(\sqrt{\frac{1}{n}}\right) \text{ and } \frac{1}{n_v}\sum_{i=1}^{n_v}\left(\ln \widehat{f}(v_i) - \ln f_0(v_i)\right) = o_p\left(\sqrt{\frac{1}{n}}\right).$$

Proof. Applying the second order Taylor expansion where $\widetilde{\theta}$ lies between θ_K^* and $\widehat{\theta}_K$, we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left(\ln \widehat{f}(x_i) - \ln f_K^*(x_i)\right) \qquad (55)\\ = & \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial f(x_i,\widehat{\theta}_K)}{\partial\theta'}}{f(x_i,\widehat{\theta}_K)}(\widehat{\theta}_K - \theta_K^*) - \frac{1}{2}(\theta_K^* - \widehat{\theta}_K)'\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial^2 f(x_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(x_i,\widetilde{\theta})}(\theta_K^* - \widehat{\theta}_K) \\ & + \frac{1}{2}(\theta_K^* - \widehat{\theta}_K)'\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial f(x_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(x_i,\widetilde{\theta})}{\partial\theta'}}{f(x_i,\widetilde{\theta})^2}(\theta_K^* - \widehat{\theta}_K). \end{aligned}$$

First, consider $\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial f(x_i,\widehat{\theta}_K)}{\partial\theta'}}{f(x_i,\widehat{\theta}_K)}(\widehat{\theta}_K - \theta_K^*) = 0$ since $\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial f(x_i,\widehat{\theta}_K)}{\partial\theta}}{f(x_i,\widehat{\theta}_K)} = 0$ from the F.O.C of (6). Next, note

$$\begin{aligned} & -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial^2 f(x_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(x_i,\widetilde{\theta})} + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial f(x_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(x_i,\widetilde{\theta})}{\partial\theta'}}{f(x_i,\widetilde{\theta})^2} \qquad (56)\\ = & -\frac{1}{n}\sum_{i=1}^{n}\frac{W^K(x_i)W^K(x_i)'}{f(x_i,\widetilde{\theta})} + 2\frac{1}{n}\sum_{i=1}^{n}\frac{(W^K(x_i)'\widetilde{\theta})^2 W^K(x_i)W^K(x_i)'}{f(x_i,\widetilde{\theta})^2} \\ = & \left(1 - 2\frac{\epsilon_0\phi(x_i)/\int_{\mathcal{X}}\phi(x)dx}{f(x_i,\widetilde{\theta})}\right)\left(\frac{f_0(x_i)}{f(x_i,\widetilde{\theta})}\right)\frac{W^K(x_i)W^K(x_i)'}{f_0(x_i)} \\ \le & \sup_{x\in\mathcal{X},\theta\in\Theta_n}\left(1 - 2\frac{\epsilon_0\phi(x)/\int_{\mathcal{X}}\phi(x)dx}{f(x,\theta)}\right)\sup_{x\in\mathcal{X}}\left(\frac{f_0(x)}{f(x,\widetilde{\theta})}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{W^K(x_i)W^K(x_i)'}{f_0(x_i)} \le C\frac{1}{n}\sum_{i=1}^{n}\frac{W^K(x_i)W^K(x_i)'}{f_0(x_i)} \end{aligned}$$

similarly with (42) noting $0 < \sup_{x\in\mathcal{X},\theta\in\Theta_n}\left(1-2\frac{\epsilon_0\phi(x)/\int_{\mathcal{X}}\phi(x)dx}{f(x,\theta)}\right) < 1$ for sufficiently small ϵ_0 and noting $\sup_{x\in\mathcal{X}}\left(\frac{f_0(x)}{f(x,\widetilde{\theta})}\right) < \infty$. This implies that the eigenvalues of (56) are bounded from above, since by Newey (1997) (noting $E\left[\frac{W^K(x_i)W^K(x_i)'}{f_0(x_i)}\right] = \int_{x\in\mathcal{X}} W^K(x)W^K(x)'dx = I_K$), we have

$$\left\|\frac{1}{n}\sum_{i=1}^n \frac{W^K(x_i)W^K(x_i)'}{f_0(x_i)} - I_K\right\| = O_p\left(\zeta(K)\sqrt{\frac{K}{n}}\right) \tag{57}$$

which implies the probability that the largest eigenvalue of $\frac{1}{n}\sum_{i=1}^n \frac{W^K(x_i)W^K(x_i)'}{f_0(x_i)}$ is smaller than 3/2 goes to one. Therefore, from (55), we conclude that

$$\begin{aligned} & \left|\frac{1}{n}\sum_{i=1}^n \left(\ln\widehat{f}(x_i) - \ln f_K^*(x_i)\right)\right| \qquad (58) \\ \leq\ & \left\|\theta_K^* - \widehat{\theta}_K\right\|^2 \lambda_{\max}\left(-\frac{1}{2}\frac{1}{n}\sum_{i=1}^n \frac{\frac{\partial^2 f(x_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(x_i,\widetilde{\theta})} + \frac{1}{2}\frac{1}{n}\sum_{i=1}^n \frac{\frac{\partial f(x_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(x_i,\widetilde{\theta})}{\partial\theta'}}{f(x_i,\widetilde{\theta})^2}\right) \\ \leq\ & O_p\left(\left\|\theta_K^* - \widehat{\theta}_K\right\|^2\right) = o_p\left(n^{-1+\alpha+2\delta}\right) \end{aligned}$$

by Lemma 2.2. Using the notation in Appendix D, consider

$$\begin{aligned} & \left|\frac{1}{n}\sum_{i=1}^n (\ln f_K^*(x_i) - \ln f_0(x_i))\right| \qquad (59) \\ =\ & \left|\widehat{Q}_n(\theta_K^*) - Q(\theta_K^*) + Q(\theta_K^*) - Q(\theta_K) + Q(\theta_K^*) - \widehat{Q}_n(\theta_K) - \left(\frac{1}{n}\sum_{i=1}^n \ln f_0(x_i) - \frac{1}{n}\sum_{i=1}^n \ln f_K(x_i)\right)\right| \\ \leq\ & \left|\widehat{Q}_n(\theta_K^*) - Q(\theta_K^*) - \left(\widehat{Q}_n(\theta_K) - Q(\theta_K)\right)\right| + \left|Q(\theta_K^*) - Q(\theta_K)\right| + \left|\frac{1}{n}\sum_{i=1}^n \ln f_0(x_i) - \frac{1}{n}\sum_{i=1}^n \ln f_K(x_i)\right|. \end{aligned}$$

Now note $\left|\widehat{Q}_n(\theta_K^*) - Q(\theta_K^*) - \left(\widehat{Q}_n(\theta_K) - Q(\theta_K)\right)\right| = o_p\left(n^{-1/2+\alpha/2+\delta}\right) O\left(K^{-s/2}\right)$ by Lemma 2.1 and Lemma D.6 and note $|Q(\theta_K^*) - Q(\theta_K)| = O\left(K^{-s}\right)$ by (40). Similarly with (37), using $|\ln(1+t)| \leq 2|t|$ in a neighborhood of $t = 0$, we also bound the last term as

$$\begin{aligned} & \left|\frac{1}{n}\sum_{i=1}^n \ln f_0(x_i) - \frac{1}{n}\sum_{i=1}^n \ln f_K(x_i)\right| \qquad (60) \\ \leq\ & \frac{1}{n}\sum_{i=1}^n \left|\ln\frac{f_0(x_i)}{f_K(x_i)}\right| \leq \frac{1}{n}\sum_{i=1}^n 2\left|\frac{f_0(x_i) - f(x_i,\theta_K)}{f(x_i,\theta_K)}\right| \\ =\ & \frac{1}{n}\sum_{i=1}^n 2\left|\frac{1}{f(x_i,\theta_K)}\right| \left|h_{f_0}(x_i)e^{-x_i^2/4} + W^K(x_i)'\theta_K\right| \left|h_{f_0}(x_i)e^{-x_i^2/4} - W^K(x_i)'\theta_K\right| \\ \leq\ & \sup_{x\in\mathcal{X}} 2\left|\frac{\sqrt{f_0(x)}}{f(x,\theta_K)}\right| \sup_{x\in\mathcal{X}}\left|h_{f_0}(x)e^{-x^2/4} - W^K(x)'\theta_K\right| \frac{1}{n}\sum_{i=1}^n\left(\left|\frac{h_{f_0}(x_i)e^{-x_i^2/4}}{\sqrt{f_0(x_i)}}\right| + \left|\frac{W^K(x_i)'\theta_K}{\sqrt{f_0(x_i)}}\right|\right). \end{aligned}$$

Note $\sup_{x\in\mathcal{X}} 2\left|\frac{\sqrt{f_0(x)}}{f(x,\theta_K)}\right| < \infty$ and $\sup_{x\in\mathcal{X}}\left|h_{f_0}(x)e^{-x^2/4} - W^K(x)'\theta_K\right| = O(K^{-s})$. The Markov inequality gives $\frac{1}{n}\sum_{i=1}^n\left(\left|\frac{h_{f_0}(x_i)e^{-x_i^2/4}}{\sqrt{f_0(x)}}\right|\right) = O_p(1)$ from (38). Finally recall $\frac{1}{n}\sum_{i=1}^n\left(\left|\frac{W^K(x_i)'\theta_K}{\sqrt{f_0(x_i)}}\right|\right) = O_p(1)$ by the

Markov inequality from (39). Thus, we conclude

$$\left|\frac{1}{n}\sum_{i=1}^{n}\ln f_0(x_i)-\frac{1}{n}\sum_{i=1}^{n}\ln f_K(x_i)\right|=O(K^{-s})$$

and hence from (59), it follows that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\ln f_K^*(x_i)-\ln f_0(x_i)\right)\right|=o_p\left(n^{-1/2+\alpha/2+\delta}\right)O\left(K^{-s/2}\right)+O(K^{-s}). \tag{61}$$

Thus, from (58) and (61), we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\ln\widehat{f}(x_i)-\ln f_0(x_i)\right)\right| &\leq \left|\frac{1}{n}\sum_{i=1}^{n}\left(\ln\widehat{f}(x_i)-\ln f_K^*(x_i)\right)\right|+\left|\frac{1}{n}\sum_{i=1}^{n}\left(\ln f_K^*(x_i)-\ln f_0(x_i)\right)\right|\\ &= o_p\left(n^{-1+\alpha+2\delta}\right)+o_p\left(n^{-1/2+\alpha/2+\delta}\right)O\left(K^{-s/2}\right)+O(K^{-s})+O(K^{-s})=o_p\left(n^{-1+\alpha+2\delta}\right)+O(K^{-s})\end{aligned}$$

which implies for $K=O\left(n^{\alpha}\right)$,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\ln\widehat{f}(x_i)-\ln f_0(x_i)\right)=o_p\left(\sqrt{\frac{1}{n}}\right)$$

with $\alpha\leq\frac{1}{2}-2\delta$ and $s>\frac{1}{2\alpha}$. Now we derive the bound for $\frac{1}{n_v}\sum\limits_{i=1}^{n_v}\left(\ln\widehat{f}(v_i)-\ln f_0(v_i)\right)$. First, we bound $\frac{1}{n_v}\sum\limits_{i=1}^{n_v}\left(\ln\widehat{f}(v_i)-\ln f_K^*(v_i)\right)$. Applying the second order Taylor expansion where $\widetilde{\theta}$ lies between θ_K^* and $\widehat{\theta}_K$, we have

$$\begin{aligned}&\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\ln\widehat{f}(v_i)-\ln f_K^*(v_i)\right)\\ =\;&\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widehat{\theta}_K)}{\partial\theta'}}{f(v_i,\widehat{\theta}_K)}(\widehat{\theta}_K-\theta_K^*)-\frac{1}{2}(\theta_K^*-\widehat{\theta}_K)'\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\frac{\partial^2 f(v_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(v_i,\widetilde{\theta})}-\frac{\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta'}}{f(v_i,\widetilde{\theta})^2}\right)(\theta_K^*-\widehat{\theta}_K).\end{aligned}$$

Define $\check{\theta}_K=\underset{\theta\in\Theta_n,f\in\mathcal{F}_n}{\operatorname{argmax}}\ \frac{1}{n_v}\sum_{i=1}^{n_v}\ln f\left(v_i,\theta\right)$ and hence $\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\check{\theta}_K)}{\partial\theta}}{f(v_i,\check{\theta}_K)}=0$, assuming $K(n_v)=K(n)$ for sufficiently large n, which can be justified since $n_v=\rho n$ for some positive constant ρ. Consider

$$\begin{aligned}\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widehat{\theta}_K)}{\partial\theta}}{f(v_i,\widehat{\theta}_K)} &= \frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widehat{\theta}_K)}{\partial\theta}}{f(v_i,\widehat{\theta}_K)}-\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widetilde{\theta}_K)}{\partial\theta}}{f(v_i,\widetilde{\theta}_K)}\\ &= 2\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\left(W^K(v_i)'\widehat{\theta}_K\right)W^K(v_i)}{f(v_i,\widehat{\theta}_K)}-2\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\left(W^K(v_i)'\widetilde{\theta}_K\right)W^K(v_i)}{f(v_i,\widetilde{\theta}_K)}\\ &= 2\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\left(\left(W^K(v_i)'\widehat{\theta}_K\right)f(v_i,\widetilde{\theta}_K)-\left(W^K(v_i)'\widetilde{\theta}_K\right)f(v_i,\widehat{\theta}_K)\right)W^K(v_i)}{f(v_i,\widehat{\theta}_K)f(v_i,\widetilde{\theta}_K)}.\end{aligned}$$

Now note

$$
\begin{aligned}
& 2\left(\left(W^K(v_i)'\widehat{\theta}_K\right) f(v_i,\check{\theta}_K) - \left(W^K(v_i)'\check{\theta}_K\right) f(v_i,\widehat{\theta}_K)\right) \\
= \; & 2\left(\left(W^K(v_i)'\widehat{\theta}_K\right)\left(W^K(v_i)'\check{\theta}_K\right) W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right) - \left(\epsilon_0\phi(v_i)/\int_{\mathcal{X}}\phi(v)dv\right) W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right)\right) \\
= \; & \left(2\left(W^K(v_i)'\widehat{\theta}_K\right)\left(W^K(v_i)'\check{\theta}_K\right) - 2\epsilon_0\phi(v_i)/\int_{\mathcal{X}}\phi(v)dv\right) W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right) \\
= \; & \left(\left(W^K(v_i)'\widehat{\theta}_K\right)^2 + \left(W^K(v_i)'\check{\theta}_K\right)^2 - \left(W^K(v_i)'\widehat{\theta}_K - W^K(v_i)'\check{\theta}_K\right)^2 - 2\epsilon_0\phi(v_i)/\int_{\mathcal{X}}\phi(v)dv\right) \\
& \times W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right) \\
= \; & \left(f(v_i,\widehat{\theta}_K) + f(v_i,\check{\theta}_K) - 4\epsilon_0\phi(v_i)/\int_{\mathcal{X}}\phi(v)dv\right) W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right) - \left(W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right)\right)^3
\end{aligned}
$$

which implies

$$
\begin{aligned}
& \left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widehat{\theta}_K)}{\partial\theta}}{f(v_i,\widehat{\theta}_K)}\right\| \qquad (62) \\
\leq \; & \left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\frac{1}{f(v_i,\widehat{\theta}_K)}+\frac{1}{f(v_i,\check{\theta}_K)}\right) W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right) W^K(v_i)\right\| \\
& + \left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\left(W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right)\right)^3 W^K(v_i)}{f(v_i,\widehat{\theta}_K)f(v_i,\check{\theta}_K)}\right\| \\
& + \left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\frac{4\epsilon_0\phi(v_i)/\int_{\mathcal{X}}\phi(v)dv}{f(v_i,\widehat{\theta}_K)f(v_i,\check{\theta}_K)}\right) W^K(v_i)'\left(\check{\theta}_K-\widehat{\theta}_K\right) W^K(v_i)\right\| \\
\leq \; & C_1\sup_{v\in\mathcal{X}}\left\|W^K(v)\right\|^2\left\|\check{\theta}_K-\widehat{\theta}_K\right\| + C_2\left(\sup_{v\in\mathcal{X}}\left\|W^K(v)\right\|\right)^4\left\|\check{\theta}_K-\widehat{\theta}_K\right\|^3
\end{aligned}
$$

where the second inequality is obtained since $f(\cdot,\cdot)$ is bounded away from zero uniformly over $\mathcal{X}$ and Θ_n and applying the Cauchy-Schwarz inequality. Note the true density of v_i is also f_0 and hence $\theta_K^* = \underset{\theta\in\Theta_n}{\operatorname{argmax}} E\left[\ln f\left(x_i,\theta\right)\right] = \underset{\theta\in\Theta_n}{\operatorname{argmax}} E\left[\ln f\left(v_i,\theta\right)\right]$

$$
\left\|\widehat{\theta}_K-\check{\theta}_K\right\| \leq \left\|\widehat{\theta}_K-\theta_K^*\right\| + \left\|\theta_K^*-\check{\theta}_K\right\| = o_p\left(n^{-1/2+\alpha/2+\delta}\right) \qquad (63)
$$

from Lemma 2.2. From (62) and (63), it follows that

$$
\begin{aligned}
& \left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widehat{\theta}_K)}{\partial\theta'}}{f(v_i,\widehat{\theta}_K)}(\widehat{\theta}_K-\theta_K^*)\right\| \leq \left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(y_i,\widehat{\theta}_K)}{\partial\theta'}}{f(y_i,\widehat{\theta}_K)}\right\|\left\|\widehat{\theta}_K-\theta_K^*\right\| \\
= \; & \left(O\left(\zeta(K)^2\right) o_p\left(n^{-1/2+\alpha/2+\delta}\right) + O\left(\zeta(K)\right)^4 o_p\left(n^{-3/2+3\alpha/2+3\delta}\right)\right) o_p\left(n^{-1/2+\alpha/2+\delta}\right) \\
= \; & O\left(\zeta(K)^2\right) o_p\left(n^{-1+\alpha+2\delta}\right).
\end{aligned}
$$

Now similarly with (56), we also obtain

$$
-\frac{1}{2}\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\frac{\frac{\partial^2 f(v_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(v_i,\widetilde{\theta})} - \frac{\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta'}}{f(v_i,\widetilde{\theta})^2}\right) \leq C\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{W^K(v_i)W^K(v_i)'}{f_0(v_i)}
$$

and hence $\lambda_{\max}\left(-\frac{1}{2}\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\frac{\frac{\partial^2 f(v_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(v_i,\widetilde{\theta})}-\frac{\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta'}}{f(v_i,\widetilde{\theta})^2}\right)\right)<\infty$ with probability approaching to one by (57) under $\frac{\zeta(K)^2K}{n}\to 0$. It follows that

$$\begin{aligned}
&\left|\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\ln\widehat{f}(v_i)-\ln f_K^*(v_i)\right)\right| \\
\leq\ &\left\|\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\frac{\partial f(v_i,\widehat{\theta}_K)}{\partial\theta'}}{f(v_i,\widehat{\theta}_K)}\right\|\left\|\widehat{\theta}_K-\theta_K^*\right\|+\left\|\theta_K^*-\widehat{\theta}_K\right\|^2\lambda_{\max}\left(-\frac{1}{2}\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\frac{\frac{\partial^2 f(v_i,\widetilde{\theta})}{\partial\theta\partial\theta'}}{f(v_i,\widetilde{\theta})}-\frac{\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta}\frac{\partial f(v_i,\widetilde{\theta})}{\partial\theta'}}{f(v_i,\widetilde{\theta})^2}\right)\right) \\
=\ &O\left(\zeta(K)^2\right)o_p\left(n^{-1+\alpha+2\delta}\right)+o_p\left(n^{-1+\alpha+2\delta}\right)=O\left(\zeta(K)^2\right)o_p\left(n^{-1+\alpha+2\delta}\right).
\end{aligned}$$

Next, a careful inspection reveals that

$$\frac{1}{n_v}\sum_{i=1}^{n_v}(\ln f_K^*(v_i)-\ln f_0(v_i))=o_p\left(n^{-1/2+\alpha/2+\delta}\right)O\left(K^{-s/2}\right)+O(K^{-s})$$

similarly with (59) noting (i) still $\left|\widehat{Q}_{n_v}(\theta_K^*)-Q(\theta_K^*)-\left(\widehat{Q}_{n_v}(\theta_K)-Q(\theta_K)\right)\right|=o_p\left(n^{-1/2+\alpha/2+\delta}\right)O\left(K^{-s/2}\right)$ and $|Q(\theta_K^*)-Q(\theta_K)|=O\left(K^{-s}\right)$ since $\underset{\theta\in\Theta_n}{\operatorname{argmax}}\,E\left[\ln f\left(v_i,\theta\right)\right]=\theta_K^*$, the density of v_i equals to f_0, and $n_v=\rho n$ and noting (ii) $\left|\frac{1}{n_v}\sum_{i=1}^{n_v}\ln f_0(v_i)-\frac{1}{n_v}\sum_{i=1}^{n_v}\ln f_K(v_i)\right|=O(K^{-s})$ similarly with (60). Therefore,

$$\begin{aligned}
&\left|\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\ln\widehat{f}(v_i)-\ln f_0(v_i)\right)\right| \\
\leq\ &\left|\frac{1}{n_v}\sum_{i=1}^{n_v}\left(\ln\widehat{f}(v_i)-\ln f_K^*(v_i)\right)\right|+\left|\frac{1}{n}\sum_{i=1}^{n}(\ln f_K^*(v_i)-\ln f(v_i))\right|=O\left(\zeta(K)^2\right)o_p\left(n^{-1+\alpha+2\delta}\right)+O(K^{-s})
\end{aligned}$$

which becomes $o_p\left(\frac{1}{\sqrt{n}}\right)$ under $\alpha\leq\frac{1}{4}-\delta$ and $s\geq\frac{1}{2\alpha}$. ■

H Proof of Lemma 3.2

Consider

$$\begin{aligned}
&\frac{1}{n-1}\sum_{i\in\mathcal{N}_x}c_i(\gamma)\ln\left(\widehat{f}(x_i)/f(x_i)\right) \\
=\ &\frac{1}{n-1}\sum_{i=1}^{n-1}c_i(\gamma)\left(\ln\widehat{f}(x_i)-\ln f(x_i)\right)+O\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x)\right) \\
=\ &\frac{1}{n-1}\sum_{i=1,odd}^{n-1}(1+\gamma)\left(\ln\widehat{f}(x_i)-\ln f(x_i)\right)+\frac{1}{n-1}\sum_{i=1,even}^{n-1}(1-\gamma)\left(\ln\widehat{f}(x_i)-\ln f(x_i)\right) \\
&+O\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x)\right).
\end{aligned}$$

Now note under $\alpha\leq\frac{1}{4}-\delta$ and $s\geq\frac{1}{2\alpha}$,

$$\sum_{i=1,odd}^{n-1}\left(\ln\widehat{f}(x_i)-\ln f(x_i)\right)=o_p(\sqrt{n})\text{ and }\sum_{i=1,even}^{n-1}\left(\ln\widehat{f}(x_i)-\ln f(x_i)\right)=o_p(\sqrt{n})$$

from Lemma G.1 since $\{x_i\}_{i:odd}$ and $\{x_i\}_{i:even}$ are subsamples of $\{x_i\}_{i=1}^n$. Also $O\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_x)\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$ under $\alpha\le\frac{1}{4}-\frac{2}{3}\delta$, $s\ge 2+\frac{1}{4\alpha}$, $\delta_f(n)=o(n^{-\frac{1}{8}})$, and $\delta_g(n)=o(n^{-\frac{1}{8}})$ as shown in Section 3.2. Therefore under $\alpha\le\frac{1}{4}-\delta$, $s\ge\max\{\frac{1}{2\alpha},2+\frac{1}{4\alpha}\}$, $\delta_f(n)=o(n^{-\frac{1}{8}})$, and $\delta_g(n)=o(n^{-\frac{1}{8}})$, we have

$$\frac{1}{n-1}\sum_{i\in\mathcal{N}_x}c_i(\gamma)\ln\left(\widehat{f}(x_i)/f(x_i)\right)=o_p\left(\frac{1}{\sqrt{n}}\right).$$

Similarly we can show that $\frac{1}{n-1}\sum_{i\in\mathcal{N}_y}c_i(\gamma)\ln\left(\widehat{g}(y_i)/g(y_i)\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$. Now we bound

$$\begin{aligned}
&\frac{1}{n-1}\sum_{i\in\mathcal{N}_y}c_i(\gamma)\ln\left(\widehat{f}(y_{i+1})/f(y_{i+1})\right)\\
=\;&\frac{1}{n-1}\sum_{i=1}^{n-1}c_i(\gamma)\left(\ln\widehat{f}(y_{i+1})-\ln f(y_{i+1})\right)+O\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_y)\right)\\
=\;&\frac{1}{n-1}\sum_{i=1,odd}^{n-1}(1+\gamma)\left(\ln\widehat{f}(y_{i+1})-\ln f(y_{i+1})\right)+\frac{1}{n-1}\sum_{i=1,even}^{n-1}(1-\gamma)\left(\ln\widehat{f}(y_{i+1})-\ln f(y_{i+1})\right)\\
&+O\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_y)\right).
\end{aligned}$$

Again note under $\alpha\le\frac{1}{4}-\delta$ and $s\ge\frac{1}{2\alpha}$,

$$\sum_{i=1,odd}^{n-1}\left(\ln\widehat{f}(y_{i+1})-\ln f(y_{i+1})\right)=o_p(\sqrt{n})\text{ and }\sum_{i=1,even}^{n-1}\left(\ln\widehat{f}(y_{i+1})-\ln f(y_{i+1})\right)=o_p(\sqrt{n})$$

from Lemma G.1 since the true density of y_i equals to f under the null ($f=g$) and $\{y_{i+1}\}_{i:odd}$ and $\{y_{i+1}\}_{i:even}$ are subsamples of $\{y_i\}_{i=1}^n$. Also note $O\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\Pr(i\notin\mathcal{N}_y)\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$ under $\alpha\le\frac{1}{4}-\frac{2}{3}\delta$, $s\ge 2+\frac{1}{4\alpha}$, $\delta_f(n)=o(n^{-\frac{1}{8}})$, and $\delta_g(n)=o(n^{-\frac{1}{8}})$ as shown in Section 3.2. Therefore under $\alpha\le\frac{1}{4}-\delta$, $s\ge\max\{\frac{1}{2\alpha},2+\frac{1}{4\alpha}\}$, $\delta_f(n)=o(n^{-\frac{1}{8}})$, and $\delta_g(n)=o(n^{-\frac{1}{8}})$, we have

$$\frac{1}{n-1}\sum_{i\in\mathcal{N}_y}c_i(\gamma)\ln\left(\widehat{f}(y_{i+1})/f(y_{i+1})\right)=o_p\left(\frac{1}{\sqrt{n}}\right).$$

Similarly we can show that $\frac{1}{n-1}\sum_{i\in\mathcal{N}_x}c_i(\gamma)\ln\left(\widehat{g}(x_{i+1})/g(x_{i+1})\right)=o_p\left(\frac{1}{\sqrt{n}}\right)$.

References

[1] Abramowitz, M. and I. A. Stegun (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York, Dover.

[2] Bickel, P.J. and M. Rosenblatt (1973), “On Some Global Measures of the Deviations of Density Function Estimates,” *Annals of Statistics*, 1, 1071-1095.

[3] Coppejans, M. and A. R. Gallant (2002), “Cross-validated SNP Density Estimates,” *Journal of Econometrics*, 110, 27-65.

[4] De Jong, R. M. (1993), "Asymptotic Theory of Expanding Parametric Space Methods and Data Dependence in Econometrics," Ph.D. Dissertation, Free University of Amsterdam, Thesis Publishers, Amsterdam, The Netherlands.

[5] Eastwood, B. J. and A. R. Gallant (1991), "Adaptive rules for seminonparametric estimators that achieve asymptotic normality," *Econometric Theory*, 7, 307-340.

[6] Fan, J., C. Zhang, and J. Zhang (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *Annals of Statistics*, 29-1, 153-193

[7] Fenton, V. M. and A. R. Gallant (1996a), "Convergence Rates of SNP Density Estimators," *Econometrica*, 64, 719-727.

[8] Fenton, V. M. and A. R. Gallant (1996b), "Qualitative and Asymptotic Performance of SNP Density Estimators," *Journal of Econometrics*, 74, 77-118.

[9] Gallant, A. R. (1987), *Nonlinear Statistical Models*, New York, Wiley.

[10] Gallant, A. R. and D. Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.

[11] Gallant, A. R. and G. Souza (1991), "On the Asymptotic Normality of Fourier Flexible Form Estimates," *Journal of Econometrics*, 50, 329-353.

[12] Gallant, A. R. and G. Tauchen (1989), "Semiparametric Estimation of Conditionally Constrained Heterogenous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091-1120.

[13] Gallant, A. R. and G. Tauchen (1989), "A Nonparametric Approach to Nonlinear Time Series Analysis: Estimation and Simulation," in D. Brillinger, P. Caines, J. Geweke, E. Parzer, M. Rosenblatt, and M. Taqqu (eds.), *New Direction in Time Series Analysis*, New York, Springer-Verlag, 71-91.

[14] Hall, P. (1984), "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Annals of Statistics*, 14, 1-16.

[15] Hansen (2004), "Nonparametric Conditional Density Estimation," Working paper, University of Wisconsin-Madison.

[16] Hoeffding, W. (1963), "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13-30.

[17] Hong, Y. and H. White (2000), "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," Discussion Paper, Department of Economics, UCSD.

[18] Kullback, S. and R.A. Leibler (1951), "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79-86.

[19] Li, Q. (1996), "Nonparametric Testing of Closeness between Two Unknown Distribution Functions," *Econometric Reviews*, 15, 261-274.

[20] Lorentz, G. (1986), *Approximation of Functions*, New York: Chelsea Publishing Company.

[21] Muir, T. A (1960), *Treatise on the Theory of Determinants*. New York: Dover.

[22] Newey, Whitney K (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.

[23] Robinson, P. M. (1991), "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437-453.

[24] Rosenblatt, M. (1975), "A Quadratic Measure of Deviation of Two-dimensional Density Estimates and a Test of Independence," *Annals of Statistics*, 3, 1-14.

[25] Ullah, A. and R. S. Singh (1989), "Estimation of a Probability Density Function with Applications to Nonparametric Inference in Econometrics," in B. Raj (eds.) *Advances in Econometrics and Modelling, Advanced Studies in Theoretical and Applied Econometrics*, Dordrecht, Kluwer Academic Publishers, 69-83.

[26] Stone, C. J. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, V.10, 1040-1053.

[27] Stone, C.J. (1990), "Large-Sample Inference for Log-Spline Models," *The Annals of Statistics*, V.18, 717-741.

[28] Su, L. and H. White (2003), "A Nonparametric Hellinger Metric Test for Conditional Independence," unpublished manuscript, UCSD.

[29] Ullah, A. (1996), "Entropy, Divergence and Distance Measures with Econometric Applications," *Journal of Statistical Planning and Inference*, 49, 137-162.

[30] White, H. and J.M. Wooldridge (1991), "Some Results on Sieve Estimation with Dependent Observations," in W.J. Barnett, J. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, New York, Cambridge University Press, 459-493.