ESTIMATING THE INTEGRAL OF A SQUARED REGRESSION FUNCTION WITH LATIN HYPERCUBE SAMPLING

by

Wei-Liem Loh
Purdue University

Technical Report #94-28

Department of Statistics
Purdue University

September 1994

◇

ESTIMATING THE INTEGRAL OF A SQUARED REGRESSION FUNCTION WITH LATIN HYPERCUBE SAMPLING[1]

By Wei-Liem Loh

Purdue University

This article is concerned with the estimation of the integral of a squared regression function using Latin hypercube sampling. A class of generalized nearest neighbour estimators $\hat{\theta}_{k,n}$ is proposed and their properties are investigated with respect to various smoothness classes of regression functions. In particular mild conditions are established which ensure that $\hat{\theta}_{k,n}$ achieves a root-n convergence rate. It is further shown that $\hat{\theta}_{k,n}$ has an asymptotic mean squared error smaller than that of any regular estimator based on an i.i.d. sample of the same size.

1 Introduction

Latin hypercube sampling was first proposed in 1979 by McKay, Conover and Beckman as an alternative to simple random (i.i.d.) sampling in computer experiments. An attractive feature of Latin hypercube sampling is that, in contrast to simple random sampling, it simultaneously stratifies on all input dimensions. More precisely, for positive integers d and n let

(i) π_k, $1 \leq k \leq d$, be random permutations of $\{1, \ldots, n\}$ each uniformly distributed over all $n!$ possible permutations,

(ii) $U_{i_1,\ldots,i_d,j}$, $1 \leq i_1, \ldots, i_d \leq n$, $1 \leq j \leq d$, be uniform $[0,1]$ random variables and

(iii) that the $U_{i_1,\ldots,i_d,j}$'s and π_k's are all stochastically independent.

A Latin hypercube sample of size n (taken from the d-dimensional unit hypercube $[0,1]^d$) is defined to be $\{X(\pi_1(i), \ldots, \pi_d(i)) : 1 \leq i \leq n\}$ where for all $1 \leq i_1, \ldots, i_d \leq n$,

$$\begin{aligned} X_j(i_1, \ldots, i_d) &= (i_j - U_{i_1,\ldots,i_d,j})/n, \quad \forall 1 \leq j \leq d, \\ X(i_1, \ldots, i_d) &= (X_1(i_1, \ldots, i_d), \ldots, X_d(i_1, \ldots, i_d))'. \end{aligned}$$

Let $f : \mathcal{R}^d \to \mathcal{R}$ be a measurable function and for each $1 \leq k \leq d$ and $x = (x_1, \ldots, x_d)' \in [0,1]^d$, we define $g_k(x_k) = \int_{[0,1]^{d-1}} f(x) \prod_{j \neq k} dx_j$. In many computer experiments, we are interested in estimating $\mu = \int_{[0,1]^d} f(x) dx$. Let

$$\hat{\mu}_n = n^{-1} \sum_{i=1}^{n} f \circ X(\pi_1(i), \ldots, \pi_d(i)).$$

Then $\hat{\mu}_n$ is an unbiased estimator for μ based on the Latin hypercube sample $\{f \circ X(\pi_1(i), \ldots, \pi_d(i)) : 1 \leq i \leq n\}$. Using an ANOVA-type decomposition for f, namely,

$$f(x) = \sum_{k=1}^{d} g_k(x_k) - (d-1)\mu + f_{rem}(x), \quad \forall x = (x_1, \ldots, x_d)' \in [0,1]^d,$$

[1]Research supported in part by NSA Grant MDA 904-93-3011.

AMS 1991 *subject classifications.* Primary 62D05; secondary 62G05, 62G20.

Key words and phrases. Integrated squared regression function, Latin hypercube sampling, nearest neighbour estimator, nonparametric information bound, rate of convergence.

Stein (1987) proved that $Var(\hat{\mu}_n) = (1/n)\int_{[0,1]^d} f_{rem}^2(x)dx + o(1/n)$ as $n \to \infty$ if $\int_{[0,1]^d} f^2(x)dx < \infty$. This reveals that Latin hypercube sampling has an additional edge over simple random sampling in that the asymptotic variance of $\hat{\mu}_n$ is always smaller than the asymptotic variance of an analogous estimator based on simple random sampling.

The problem of estimating $\int_{[0,1]^d} f_{rem}^2(x)dx$, using Latin hypercube sampling, was studied by Owen (1992). Observing that

$$\int_{[0,1]^d} f_{rem}^2(x)dx = \int_{[0,1]^d} f^2(x)dx + (d-1)\mu^2 - \sum_{k=1}^{d}\int_0^1 g_k^2(x_k)dx_k,$$

he proposed a class of nearest neighbour estimators for $\theta_k = \int_0^1 g_k^2(t)dt$ and essentially proved that these estimators are $n^{1/2}$ consistent provided the following Lipschitz condition holds: there exists a constant M such that $|g_k(s) - g_k(t)| \leq M|s-t|$ for all $s, t \in [0,1]$.

In this paper we undertake a more detailed study of the estimation of θ_k under weaker smoothness assumptions on g_k using observations

$$\{X_k(\pi_1(i), \ldots, \pi_d(i)), f \circ X(\pi_1(i), \ldots, \pi_d(i)) : 1 \leq i \leq n\} \tag{1}$$

which are obtained via Latin hypercube sampling. In Section 2, we propose a class of generalized nearest neighbour estimators $\hat{\theta}_{k,n}$ for θ_k. Upper bounds on their convergence rates are computed over various smoothness classes of regression functions. In particular, mild conditions are obtained which ensure that $\hat{\theta}_{k,n}$ achieves a $n^{1/2}$ convergence rate (with explicit constants) as well as its asymptotic normality.

As another yardstick to gauge the performance of $\hat{\theta}_{k,n}$, Section 3 investigates the estimation of θ_k using simple random sampling. The nonparametric information bound [in the sense of Stein (1956) and Levit (1974)] for the estimation of θ_k is determined and a consequence of which is that $\hat{\theta}_{k,n}$ (under Latin hypercube sampling) possesses a smaller asymptotic mean squared error than that of *any* regular estimator for θ_k based on a simple random sample of the same size. The Appendix contains proofs of a number of technical results that are needed in Section 2.

REMARK. A related problem of estimating integrated squared density derivatives with simple random sampling was studied by Hall and Marron (1987) and Bickel and Ritov (1988). They proposed a number of estimators based on the method of kernels. In particular Bickel and Ritov showed that the convergence rate of their estimators is optimal given the amount of smoothness assumed and that they achieve the information bound when estimation at an $n^{-1/2}$ rate is possible. However due to the presence of boundaries in the current problem, it is not clear whether their technique can be used to construct similar kernel estimators for θ_k which possess convergence rates comparable to those reported in this paper.

2 A class of generalized nearest neighbour estimators

For simplicity, we shall from now on write $X(\pi_1(i), \ldots, \pi_d(i))$ as $X(\bar{\pi}(i))$, etc., and without loss of generality assume that $(i-1)/n \leq X_k(\bar{\pi}(i)) < i/n$ for all $1 \leq i \leq n$. For some positive integer m_n, let $\{w_{i,j}(m_n), \tilde{w}_{i,j}(m_n) : 1 \leq i, j \leq n\}$ be a sequence of nonnegative constants such that

$$w_{i,j}(m_n) = \tilde{w}_{i,j}(m_n) = 0, \quad \text{if } i = j \text{ or } |j-i| > m_n, \tag{2}$$

$$\sum_{j=1}^{n} w_{i,j}(m_n) = 1, \quad \forall 1 \leq i \leq n, \tag{3}$$

$$\sum_{i=1}^{n} w_{i,j}(m_n) = \begin{cases} 1, & \forall m_n+1 \leq j \leq n-m_n, \\ O(1), & \text{uniformly in } j \text{ otherwise}, \end{cases} \tag{4}$$

$$\sup\{\tilde{w}_{i,j}(m_n) : 1 \leq i,j \leq n\} = o(m_n^{-1/2}), \quad \text{as } n \to \infty, \tag{5}$$

$$\tilde{w}_{i,j}(m_n) \sum_{j_1 \neq j} \tilde{w}_{i,j_1}(m_n) = w_{i,j}(m_n), \quad \forall 1 \leq i,j \leq n. \tag{6}$$

EXAMPLE. Let $m_n > 1$ and $c_i(m_n)$ denote the cardinality of the set $\{j : 0 < |j-i| \leq m_n, 1 \leq j \leq n\}$. For $1 \leq i,j \leq n$, define

$$w^*_{i,j}(m_n) = \begin{cases} c_i^{-1}(m_n), & \text{if } 0 < |j-i| \leq m_n, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\tilde{w}^*_{i,j}(m_n) = \begin{cases} [c_i(m_n)(c_i(m_n)-1)]^{-1/2}, & \text{if } 0 < |j-i| \leq m_n, \\ 0, & \text{otherwise}. \end{cases}$$

Then the sequence $\{w^*_{i,j}(m_n), \tilde{w}^*_{i,j}(m_n) : 1 \leq i,j \leq n\}$ satisfies the conditions (2)–(6).

We propose the following class of generalized nearest neighbour estimators $\hat{\theta}_{k,n}$ for θ_k. Define

$$\begin{aligned} \hat{\theta}_{k,n} &= (2/n)\sum_{i=1}^{n}\{f \circ X(\bar{\pi}(i)) \sum_{j=1}^{n} w_{i,j}(m_n) f \circ X(\bar{\pi}(j)) \\ &\quad - \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^{n} \tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) f \circ X(\bar{\pi}(j_1)) f \circ X(\bar{\pi}(j_2))\}, \end{aligned} \tag{7}$$

where m_n is some suitably chosen integer which tends to infinity as $n \to \infty$.

Due to the stratified nature of $\{X_k(\bar{\pi}(i)) : 1 \leq i \leq n\}$, this class of estimators is especially suited to the present problem. Roughly speaking, these estimators behave like kernel estimators at the interior points of $[0,1]$ while allowing simple modifications to be made to them when estimating points near the boundaries; thus safeguarding against increased biases at the boundaries.

For robustness reasons, we shall investigate the mean squared error of $\hat{\theta}_{k,n}$ with respect to the following Lipschitz-type classes of functions. For constants $0 < \alpha \leq 1$ and $M > 0$, we define $\mathcal{F}_{k,\alpha,M}$ to be the set of all functions $f : [0,1]^d \to \mathcal{R}$ such that $\sup\{\int_{[0,1]^{d-1}} f^4(x) \prod_{j \neq k} dx_j : 0 \leq x_k \leq 1\} \leq M$ and that $|g_k(s) - g_k(t)| \leq M|s-t|^\alpha$ whenever $0 \leq s,t \leq 1$. The following is the main result of this article.

Theorem 1 *Let $1 \leq k \leq d$ and $\hat{\theta}_{k,n}$ be as in (7). Then we have*

$$\begin{aligned} E(\hat{\theta}_{k,n} - \theta_k)^2 &= 4n^{-1}\{\int_{[0,1]^d} f^2(x) g_k^2(x_k) dx - \int_0^1 g_k^4(x_k) dx_k - \sum_{j \neq k} \int_0^1 R^2_{k,j,n}(x_j) dx_j\} \\ &\quad + O((m_n/n)^{4\alpha}) + o(1/n), \end{aligned}$$

as $m_n \to \infty$ and $m_n/n \to 0$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$ where for $j \neq k$ and $0 \leq x_j \leq 1$,

$$\begin{aligned} R_{k,j,n}(x_j) &= \sum_{i=1}^{n}\{\sum_{j_1=1}^{n} n w_{i,j_1}(m_n) \int_{(i-1)/n}^{i/n} \int_{(j_1-1)/n}^{j_1/n} [\tilde{f}_{k,1,-j}(x_j,x_k,y_k) + \tilde{f}_{k,2,-j}(x_j,x_k,y_k)] dx_k dy_k \\ &\quad - \sum_{j_2=1}^{n-1} \sum_{j_3=j_2+1}^{n} n\tilde{w}_{i,j_2}(m_n)\tilde{w}_{i,j_3}(m_n) \\ &\quad \times \int_{(j_2-1)/n}^{j_2/n} \int_{(j_3-1)/n}^{j_3/n} [\tilde{f}_{k,1,-j}(x_j,x_k,y_k) + \tilde{f}_{k,2,-j}(x_j,x_k,y_k)] dx_k dy_k\}, \end{aligned}$$

with

$$\tilde{f}_{k,1,-j}(x_j, x_k, y_k) = g_k(y_k)[\int_{[0,1]^{d-2}} f(x) \prod_{l\neq j,k} dx_l - g_k(x_k)],$$
$$\tilde{f}_{k,2,-j}(y_j, x_k, y_k) = g_k(x_k)[\int_{[0,1]^{d-2}} f(y) \prod_{l\neq j,k} dy_l - g_k(y_k)].$$

The following corollary shows that $\hat{\theta}_{k,n}$ is capable of estimating θ_k at an $n^{-1/2}$ rate (with explicit constants) under mild conditions.

Corollary 1 *Under the conditions of Theorem 1, if* $1/4 < \alpha \leq 1$ *and by choosing* $m_n = o(n^{(4\alpha-1)/4\alpha})$, *we have*

$$E(\hat{\theta}_{k,n} - \theta_k)^2 = 4n^{-1}\{\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx - \int_0^1 g_k^4(x_k)dx_k - \sum_{j\neq k}\int_0^1 R_{k,j,n}^2(x_j)dx_j\} + o(1/n),$$

as $m_n \to \infty$ *uniformly over* $f \in \mathcal{F}_{k,\alpha,M}$.

The next result gives sufficient conditions for $\hat{\theta}_{k,n}$ to be asymptotically normal.

Theorem 2 *Under the conditions of Theorem 1, if* $1/4 < \alpha \leq 1$ *and choosing* $m_n = o(n^{\min\{(4\alpha-1)/4\alpha, 1/2\}})$, *we have for each* $f \in \mathcal{F}_{k,\alpha,M}$, $n^{1/2}\sigma_{k,n}^{-1}(\hat{\theta}_{k,n} - \theta_k)$ *converges in distribution to* $N(0,1)$, *as* $m_n \to \infty$ *if*

$$\sigma_{k,n}^2 = 4\{\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx - \int_0^1 g_k^4(x_k)dx_k - \sum_{j\neq k}\int_0^1 R_{k,j,n}^2(x_j)dx_j\}$$

and $\liminf_{m_n\to\infty} \sigma_{k,n}^2 > 0$.

REMARK. We observe that if f is continuous, say, then under the assumptions of Theorem 1 we have as $m_n \to \infty$,

$$\int_0^1 R_{k,j,n}^2(x_j)dx_j = \int_0^1 \{\int_{[0,1]^{d-1}} g_k(x_k)[f(x) - g_k(x_k)]\prod_{l\neq j} dx_l\}^2 dx_j + o(1), \quad \forall j \neq k.$$

3 On the nonparametric information bound

In this section we shall study the problem of estimating θ_k using a simple random sample. The hope is that this would give us another yardstick for gauging the quality of the results of the previous section. More precisely, let $X(1), \ldots, X(n)$ be an i.i.d. sample where $X(1) = (X_1(1), \ldots, X_d(1))'$ is uniformly distributed on $[0,1]^d$. The aim of this section is to compute the nonparametric information bound for the estimation of θ_k using the observations

$$\{X_k(i), f \circ X(i) : 1 \leq i \leq n\}. \tag{8}$$

For simplicity, we write $y = f(x)$ for all $x \in [0,1]^d$ and define $p(x_k, y)$ to be the joint probability density function of $X_k(1)$ and $Y = f \circ X(1)$.

Proposition 1 *The linear space of all score functions of regular parametric submodels through $p(x_k, y)$, namely the tangent space, is the space of all L_2-integrable functions [w.r.t. the measure $p(x_k, y)dx_k dy$] which are orthogonal to L_2-integrable functions which are functions only of x_k (and not y).*

The proof of Proposition 1 follows from the observation that the marginal distribution of $X_k(1)$ is known and a detailed proof is similar to that given in Bickel, Ritov and Wellner (1991), page 1328.

We further observe that the functional θ_k has pathwise derivative

$$\dot{\theta}_k(h) = 2\int_{-\infty}^{\infty}\int_0^1 yE(Y|X_k(1) = x_k)h(x_k, y)p(x_k, y)dx_k dy,$$

for all h in the tangent space. Hence it follows from Proposition 1 that the canonical gradient (or the adjoint map) is equal to $G(x_k, y) = 2[yE(Y|X_k(1) = x_k) - E^2(Y|X_k(1) = x_k)]$ for all $x_k \in [0,1], y \in \mathcal{R}$. This implies that the nonparametric information bound for the estimation of θ_k is $Var\{G(X_k(1), Y)\} = 4[\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx - \int_0^1 g_k^4(x_k)dx_k]$. Thus we have proved

Theorem 3 *Let $\tilde{\theta}_{k,n}$ be a regular estimator for θ_k based on observations as in (8). Then*

$$\liminf_{n\to\infty} nE(\tilde{\theta}_{k,n} - \theta_k)^2 \geq 4\{\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx - \int_0^1 g_k^4(x_k)dx_k\}.$$

Theorems 1 and 3 reveal that $\hat{\theta}_{k,n}$ under Latin hypercube sampling possesses an asymptotic mean squared error smaller than that of any estimator for θ_k based on a simple random sample of the same size.

4 Appendix

PROOF OF THEOREM 1. First we compute the bias of $\hat{\theta}_{k,n}$. Define for $0 \leq s, t < 1$,

$$\delta_n(s,t) = \begin{cases} 1 & \text{if } \lfloor ns \rfloor = \lfloor nt \rfloor, \\ 0 & \text{otherwise,} \end{cases}$$

where $\lfloor t \rfloor$ denotes the greatest integer less than or equal to t. Since $(i-1)/n \leq X_k(\bar{\pi}(i)) < i/n$ for all $1 \leq i \leq n$, we observe that

$$\begin{aligned} & E(2/n)\sum_{i=1}^n f \circ X(\bar{\pi}(i)) \sum_{j=1}^n w_{i,j}(m_n) f \circ X(\bar{\pi}(j)) \\ = \; & \frac{2n^d}{(n-1)^{d-1}} \sum_{i=1}^n \int_{[0,1]^{2d-2}} \int_{(i-1)/n}^{i/n} f(x) \sum_{j=1}^n \int_{(j-1)/n}^{j/n} w_{i,j}(m_n) f(y) dy_k dx_k \prod_{l \neq k} [1 - \delta_n(x_l, y_l)] dx_l dy_l \\ (9) = \; & 2\sum_{i=1}^n \int_{(i-1)/n}^{i/n} g_k(x_k) \sum_{j=1}^n \int_{(j-1)/n}^{j/n} n w_{i,j}(m_n) g_k(y_k) dy_k dx_k + O(n^{-1}), \end{aligned}$$

and similarly

$$\begin{aligned} & En^{-1}\sum_{i=1}^n \sum_{j_1=1}^n \sum_{j_2 \neq j_1} \tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) f \circ X(\bar{\pi}(j_1)) f \circ X(\bar{\pi}(j_2)) \\ = \; & \sum_{i=1}^n \sum_{j_1=1}^n \sum_{j_2 \neq j_1} \int_{(i-1)/n}^{i/n} \int_{(j_1-1)/n}^{j_1/n} \int_{(j_2-1)/n}^{j_2/n} n^2 \tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) \\ (10) \quad & \times g_k(y_k) g_k(z_k) dy_k dz_k dx_k + O(n^{-1}), \end{aligned}$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$. Now it follows from (3), (6), (7), (9) and (10) that

$$\begin{aligned} E\hat{\theta}_{k,n} - \theta_k &= -\sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2\neq j_1}\int_{(i-1)/n}^{i/n}\int_{(j_1-1)/n}^{j_1/n}\int_{(j_2-1)/n}^{j_2/n} n^2\tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) \\ &\quad \times[g_k(y_k)-g_k(x_k)][g_k(z_k)-g_k(x_k)]dy_k dz_k dx_k + O(n^{-1}), \end{aligned}$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$. This "factorization" is crucial in further reducing the bias by another order. Using (3), (6) and the Lipschitz condition of $f \in \mathcal{F}_{k,\alpha,M}$, we have

$$\sup\{|E\hat{\theta}_{k,n} - \theta_k| : f \in \mathcal{F}_{k,\alpha,M}\} = O((m_n/n)^{2\alpha} + (1/n)), \tag{11}$$

as $n \to \infty$. Next we shall compute the variance of $\hat{\theta}_{k,n}$. Observe that

$$\begin{aligned} E(\hat{\theta}_{k,n} - E\hat{\theta}_{k,n})^2 &= \frac{4}{n^2}Var\{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(m_n) f\circ X(\bar{\pi}(i)) f\circ X(\bar{\pi}(j))\} \\ &\quad +\frac{1}{n^2}Var\{\sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2\neq j_1}\tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) f\circ X(\bar{\pi}(j_1)) f\circ X(\bar{\pi}(j_2))\} \\ &\quad -\frac{4}{n^2}Cov\{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(m_n) f\circ X(\bar{\pi}(i)) f\circ X(\bar{\pi}(j)), \\ &\quad \sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2\neq j_1}\tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) f\circ X(\bar{\pi}(j_1)) f\circ X(\bar{\pi}(j_2))\}. \end{aligned}$$

Now it follows from Lemmas 2, 3 and 4 below that

$$\begin{aligned} & E(\hat{\theta}_{k,n} - E\hat{\theta}_{k,n})^2 \\ = \;& 4n^{-1}\{\int_{[0,1]^d} f^2(x) g_k^2(x_k)dx - \int_0^1 g_k^4(x_k)dx_k\} - n^{-1}\sum_{j\neq k}\int_0^1\{\{2\sum_{i_1,j_1=1}^{n} n w_{i_1,j_1}(m_n) \\ & \times\int_{(i_1-1)/n}^{i_1/n}\int_{(j_1-1)/n}^{j_1/n}[\tilde{f}_{k,1,-j}(x_j,x_k,y_k)+\tilde{f}_{k,2,-j}(x_j,x_k,y_k)]dx_k dy_k\} \\ & -\{\sum_{i,j_1=1}^{n}\sum_{j_2\neq j_1} n\tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) \\ & \times\int_{(j_1-1)/n}^{j_1/n}\int_{(j_2-1)/n}^{j_2/n}[\tilde{f}_{k,1,-j}(x_j,x_k,y_k)+\tilde{f}_{k,2,-j}(x_j,x_k,y_k)]dx_k dy_k\}\}^2 dx_j + o(1/n), \end{aligned} \tag{12}$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$. This proves Theorem 1 since the mean squared error of $\hat{\theta}_{k,n}$ is equal to the sum of its variance and squared bias. □

Lemma 1 *With the notation of Theorem 1, if i_1, i_2, j_1, j_2 are all distinct integers then*

$$\begin{aligned} & Cov\{f\circ X(\bar{\pi}(i_1)) f\circ X(\bar{\pi}(j_1)), f\circ X(\bar{\pi}(i_2)) f\circ X(\bar{\pi}(j_2))\} \\ = \;& -n^3\sum_{j\neq k}\int_0^1\{\int_{(i_1-1)/n}^{i_1/n}\int_{(j_1-1)/n}^{j_1/n}[\tilde{f}_{k,1,-j}(x_j,x_k,y_k)+\tilde{f}_{k,2,-j}(x_j,x_k,y_k)]dx_k dy_k\} \\ & \times\{\int_{(i_2-1)/n}^{i_2/n}\int_{(j_2-1)/n}^{j_2/n}[\tilde{f}_{k,1,-j}(x_j,x_k,y_k)+\tilde{f}_{k,2,-j}(x_j,x_k,y_k)]dx_k dy_k\}dx_j + O(n^{-2}), \end{aligned} \tag{13}$$

and

$$
\begin{aligned}
& Cov\{f \circ X(\bar\pi(i_1)) f \circ X(\bar\pi(j_1)), f \circ X(\bar\pi(i_1)) f \circ X(\bar\pi(j_2))\} \\
= \; & n^3 \int_{[0,1]^{d-1}} \int_{(i_1-1)/n}^{i_1/n} \int_{(j_1-1)/n}^{j_1/n} \int_{(j_2-1)/n}^{j_2/n} f^2(x) g_k(y_k) g_k(z_k) dy_k dz_k dx_k \prod_{l \neq k} dx_l \\
& -[n^2 \int_{(i_1-1)/n}^{i_1/n} \int_{(j_1-1)/n}^{j_1/n} g_k(x_k) g_k(y_k) dy_k dx_k] \\
& \times [n^2 \int_{(i_1-1)/n}^{i_1/n} \int_{(j_2-1)/n}^{j_2/n} g_k(x'_k) g_k(y'_k) dy'_k dx'_k] + O(1/n), \qquad (14)
\end{aligned}
$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$.

PROOF. First we observe that for $i \neq j$,

$$
\begin{aligned}
& E f \circ X(\bar\pi(i)) f \circ X(\bar\pi(j)) \\
= \; & \frac{n^{d+1}}{(n-1)^{d-1}} \int_{[0,1]^{2d-2}} \int_{(i-1)/n}^{i/n} \int_{(j-1)/n}^{j/n} f(x) f(y) dx_k dy_k \prod_{l \neq k} [1 - \delta_n(x_l, y_l)] dx_l dy_l \\
= \; & n^2 \int_{(i-1)/n}^{i/n} \int_{(j-1)/n}^{j/n} g_k(x_k) g_k(y_k) dx_k dy_k + O(1/n),
\end{aligned}
$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$. Hence using the "orthogonal" decomposition

$$
f(x) f(y) = g_k(x_k) g_k(y_k) + \sum_{j \neq k} [\tilde f_{k,1,-j}(x_j, x_k, y_k) + \tilde f_{k,2,-j}(y_j, x_k, y_k)] + \tilde f_{k,rem}(x, y),
$$

we observe that the l.h.s. of (13) is equal to

$$
\begin{aligned}
& -n^3 \sum_{j \neq k} \int_0^1 \{ \int_{(i_1-1)/n}^{i_1/n} \int_{(j_1-1)/n}^{j_1/n} [\tilde f_{k,1,-j}(x_j, x_k, y_k) + \tilde f_{k,2,-j}(x_j, x_k, y_k)] dx_k dy_k \} \\
& \times \{ \int_{(i_2-1)/n}^{i_2/n} \int_{(j_2-1)/n}^{j_2/n} [\tilde f_{k,1,-j}(x_j, x'_k, y'_k) + \tilde f_{k,2,-j}(x_j, x'_k, y'_k)] dx'_k dy'_k \} dx_j + O(n^{-2}),
\end{aligned}
$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$. The proof of the second statement of the lemma is similar and is omitted. □

Lemma 2 *With the notation and assumptions of Theorem 1,*

$$
\begin{aligned}
& Var\{\sum_{i=1}^n \sum_{j=1}^n w_{i,j}(m_n) f \circ X(\bar\pi(i)) f \circ X(\bar\pi(j))\} \\
= \; & 4n\{ \int_{[0,1]^d} f^2(x) g_k^2(x_k) dx - \int_0^1 g_k^4(x_k) dx_k \} - n \sum_{j \neq k} \int_0^1 \{ \sum_{i_1, j_1 = 1}^n n w_{i_1,j_1}(m_n) \\
& \times \int_{(i_1-1)/n}^{i_1/n} \int_{(j_1-1)/n}^{j_1/n} [\tilde f_{k,1,-j}(x_j, x_k, y_k) + \tilde f_{k,2,-j}(x_j, x_k, y_k)] dx_k dy_k \}^2 dx_j + o(n), \qquad (15)
\end{aligned}
$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$.

PROOF. We observe that the l.h.s. of (15) is equal to

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}^2(m_n)Var\{f\circ X(\bar\pi(i))f\circ X(\bar\pi(j))\}$$
$$+\sum_{i,j_1,j_2}{}^{*}[w_{i,j_1}(m_n)w_{i,j_2}(m_n)+w_{j_1,i}(m_n)w_{j_2,i}(m_n)+w_{i,j_1}(m_n)w_{j_2,i}(m_n)+w_{j_1,i}(m_n)w_{i,j_2}(m_n)]$$
$$\times Cov\{f\circ X(\bar\pi(i))f\circ X(\bar\pi(j_1)), f\circ X(\bar\pi(i))f\circ X(\bar\pi(j_2))\}$$
$$(16)\ +\sum{}^{**} w_{i_1,j_1}(m_n)w_{i_2,j_2}(m_n)Cov\{f\circ X(\bar\pi(i_1))f\circ X(\bar\pi(j_1)), f\circ X(\bar\pi(i_2))f\circ X(\bar\pi(j_2))\},$$

where $\sum_{i,j_1,j_2}^{*}$ denotes summation over all distinct values of $1\le i,j_1,j_2\le n$ and $\sum^{**}$ denotes summation over all distinct values of $1\le i_1,i_2,j_1,j_2\le n$. From (3) and (5), we observe that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}^2(m_n)Var\{f\circ X(\bar\pi(i))f\circ X(\bar\pi(j))\}=o(n),$$

as $n\to\infty$ uniformly over $f\in\mathcal{F}_{k,\alpha,M}$. Next using (3), (4), (14) and the Lipschitz condition on g_k, we observe that the second term in (16) is equal to

$$4n\{\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx-\int_0^1 g_k^4(x_k)dx_k\}+o(n),$$

as $n\to\infty$ uniformly over $f\in\mathcal{F}_{k,\alpha,M}$. Finally we observe from (13) that the third term in (16) is equal to

$$-n\sum_{j\neq k}\int_0^1\{\sum_{i_1,j_1=1}^{n} nw_{i_1,j_1}(m_n)$$
$$\times\int_{(i_1-1)/n}^{i_1/n}\int_{(j_1-1)/n}^{j_1/n}[\tilde f_{k,1,-j}(x_j,x_k,y_k)+\tilde f_{k,2,-j}(x_j,x_k,y_k)]dx_kdy_k\}^2dx_j+o(n),$$

as $n\to\infty$ uniformly over $f\in\mathcal{F}_{k,\alpha,M}$. This proves Lemma 2. □

Lemma 3 *With the notation and assumptions of Theorem 1,*

$$Var\{\sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2\neq j_1}\tilde w_{i,j_1}(m_n)\tilde w_{i,j_2}(m_n)f\circ X(\bar\pi(j_1))f\circ X(\bar\pi(j_2))\}$$
$$=\ 4n\{\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx-\int_0^1 g_k^4(x_k)dx_k\}-n\sum_{j\neq k}\int_0^1\{\sum_{i,j_1=1}^{n}\sum_{j_2\neq j_1} n\tilde w_{i,j_1}(m_n)\tilde w_{i,j_2}(m_n)$$
$$\times\int_{(j_1-1)/n}^{j_1/n}\int_{(j_2-1)/n}^{j_2/n}[\tilde f_{k,1,-j}(x_j,x_k,y_k)+\tilde f_{k,2,-j}(x_j,x_k,y_k)]dx_kdy_k\}^2dx_j+o(n),$$

as $n\to\infty$ uniformly over $f\in\mathcal{F}_{k,\alpha,M}$.

Lemma 4 *With the notation and assumptions of Theorem 1,*

$$Cov\{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}(m_n)f\circ X(\bar\pi(i))f\circ X(\bar\pi(j)),$$

$$\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2\neq j_1}\tilde{w}_{i_1,j_1}(m_n)\tilde{w}_{i_1,j_2}(m_n)f\circ X(\bar{\pi}(j_1))f\circ X(\bar{\pi}(j_2))\}$$
$$\begin{aligned}
= \;& 4n\{\int_{[0,1]^d} f^2(x)g_k^2(x_k)dx - \int_0^1 g_k^4(x_k)dx_k\} - n\sum_{j\neq k}\int_0^1\{\sum_{i_1,j_1=1}^{n} nw_{i_1,j_1}(m_n) \\
&\times\int_{(i_1-1)/n}^{i_1/n}\int_{(j_1-1)/n}^{j_1/n}[\tilde{f}_{k,1,-j}(x_j,x_k,y_k)+\tilde{f}_{k,2,-j}(x_j,x_k,y_k)]dx_kdy_k\} \\
&\times\{\sum_{i,j_1=1}^{n}\sum_{j_2\neq j_1} n\tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n) \\
&\times\int_{(j_1-1)/n}^{j_1/n}\int_{(j_2-1)/n}^{j_2/n}[\tilde{f}_{k,1,-j}(x_j,x_k',y_k')+\tilde{f}_{k,2,-j}(x_j,x_k',y_k')]dx_k'dy_k'\}dx_j + o(n),
\end{aligned}$$

as $n \to \infty$ uniformly over $f \in \mathcal{F}_{k,\alpha,M}$.

The proofs of Lemmas 3 and 4 are similar to that of Lemma 2 and are omitted.

PROOF OF THEOREM 2. We observe from (11) and (12) that it suffices to show that $n^{1/2}(\hat{\theta}_{k,n} - E\hat{\theta}_{k,n})$ is asymptotically normal. For $1 \leq i \leq n$, define $\mu_{k,n}(i) = n\int_{(i-1)/n}^{i/n} g_k(x_k)dx_k$. Using the decomposition

$$\begin{aligned}
& f(x)f(y) - \mu_{k,n}(i)\mu_{k,n}(j) \\
= \;& \mu_{k,n}(i)[f(x)-\mu_{k,n}(j)] + \mu_{k,n}(j)[f(y)-\mu_{k,n}(i)] + [f(x)-\mu_{k,n}(j)][f(y)-\mu_{k,n}(i)],
\end{aligned}$$

we obtain

$$\begin{aligned}
n^{1/2}(\hat{\theta}_{k,n} - E\hat{\theta}_{k,n}) \;=\;& 4n^{-1/2}\sum_{i=1}^{n}[f\circ X(\bar{\pi}(i)) - \mu_{k,n}(i)]\sum_{j=1}^{n} w_{i,j}(m_n)\mu_{k,n}(j) \\
& -2n^{-1/2}\sum_{i=1}^{n}[f\circ X(\bar{\pi}(i)) - \mu_{k,n}(i)]\sum_{j_1=1}^{n}\sum_{j\neq i}\tilde{w}_{j_1,i}(m_n)\tilde{w}_{j_1,j}(m_n)\mu_{k,n}(j) \\
& +\Delta_{k,n} + O(n^{-1/2}),
\end{aligned}$$

as $n \to \infty$ where

$$\begin{aligned}
\Delta_{k,n} \;=\;& 2n^{-1/2}\sum_{i=1}^{n}\{\sum_{j=1}^{n} w_{i,j}(m_n)[f\circ X(\bar{\pi}(i)) - \mu_{k,n}(i)][f\circ X(\bar{\pi}(j)) - \mu_{k,n}(j)] \\
& -\sum_{j_1=1}^{n-1}\sum_{j_2=j_1+1}^{n}\tilde{w}_{i,j_1}(m_n)\tilde{w}_{i,j_2}(m_n)[f\circ X(\bar{\pi}(j_1)) - \mu_{k,n}(j_1)][f\circ X(\bar{\pi}(j_2)) - \mu_{k,n}(j_2)]\}.
\end{aligned}$$

As in the proofs of Lemmas 2 to 4, we have $E\Delta_{k,n}^2 = o(1)$ and, from Markov's inequality, $\Delta_{k,n} = o_p(1)$ as $m_n \to \infty$. From (3), (4) and (6), and making repeated use of Markov's inequality to get bounds for the remainder terms, we have for $f \in \mathcal{F}_{k,\alpha,M}$

$$n^{1/2}(\hat{\theta}_{k,n} - E\hat{\theta}_{k,n}) = 2n^{-1/2}\sum_{i=1}^{n}[f\circ X(\bar{\pi}(i)) - \mu_{k,n}(i)]\mu_{k,n}(i) + o_p(1), \tag{17}$$

as $m_n \to \infty$. From Theorem 1 of Loh (1993), we note that the first term on the r.h.s. of (17) is asymptotically normal and Theorem 2 is proved. □

References

[1] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya* **50** 381-393.

[2] BICKEL, P. J., RITOV, Y. and WELLNER, J. A. (1991). Efficient estimation of linear functionals of a probability measure P with known marginal distributions. *Ann. Statist.* **19** 1316-1346.

[3] HALL, P. and MARRON, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109-115.

[4] LEVIT, B. Y. (1974). On optimality of some statistical estimators. In *Proc. Prague Symp. Asymp. Statist.* (J. Hájek, ed.) **2** 215-238. Univ. Karlova, Prague.

[5] LOH, W. L. (1993). On Latin hypercube sampling. Tech. Report No. 93-52, Purdue University.

[6] MCKAY, M. D., CONOVER, W. J. and BECKMAN, R. J. (1979). A comparison of three methods for selecting values of output variables in the analysis of output from a computer code. *Technometrics* **21** 239-245.

[7] OWEN, A. B. (1992). A central limit theorem for Latin hypercube sampling. *J. R. Statist. Soc.* **54** 541-551.

[8] STEIN, C. M. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187-195. Univ. California Press.

[9] STEIN, M. L. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29** 143-151.