Uniform Bound on Normal Approximation of Latin Hypercube Sampling

N. Rerkruthairat & K. Neammanee

Department of Mathematics, Faculty of Science

Chulalongkorn University, Bangkok 10330, Thailand

E-mail: Kritsana.N@chula.ac.th

Received: October 28, 2010 Accepted: November, 2012 doi:10.5539/jmr.v3n2p98

The research is financed by Commission on Higher Education (Sponsoring information)

Abstract

Loh (Loh, W.L, 1996b) established a Berry-Esseen type bound for W, the random variable based on a latin hypercube sampling, to the standard normal distribution. He used an inductive approach of Stein's method to give the rate of convergence $\frac{C_d}{\sqrt{n}}$ without the value of C_d. In this article, we use a concentration inequality approach of Stein's method to obtain a constant C_d.

Keywords: Latin hypercube sampling, Stein's method, Uniform bound, Berry-Esseen theorem, Concentration inequality

1. Introduction

A latin hypercube sampling (LHS) was introduced by McKay, Beckman and Conover in 1978 (McKay, M.D., 1979) as a tool to improve the efficiency of different important sampling method. After the original paper appeared, LHS has been widely used in many computer experiments. For example, it is a way to choose the points to compute the integral

$$\mu = \int_{[0,1]^d} f(x)dx,$$

where f is a measurable function from $[0, 1]^d$ to $\mathbb{R}$. Approximating this integral is equivalent to finding $\mu = E(f(X))$, where X is a random vector uniformly distributed on a unit hypercube $[0, 1]^d$.

For positive integers n and d, $d \geq 2$, a latin hypercube sample of size n (taken from the d-dimensional hypercube $[0, 1]^d$) is defined to be $\{X(\eta_1(i), \eta_2(i), ..., \eta_d(i)) : 1 \leq i \leq n\}$, where

1. for all $1 \leq i_1, ..., i_d \leq n,\ 1 \leq j \leq d$,

$$X_j(i_1, ..., i_d) = (i_j - U_{i_1,...,i_d,j})/n, \text{ and } X(i_1, ..., i_d) = (X_1(i_1, ..., i_d), ..., X_d(i_1, ..., i_d));$$

2. $\eta_k = (\eta_k(1), \eta_k(2), \ldots, \eta_k(n)),\ 1 \leq k \leq d$, are random permutations of $\{1, ..., n\}$ each uniformly distributed over all the $n!$ possible permutations;

3. $U_{i_1,...,i_d,j},\ 1 \leq i_1, ..., i_d \leq n,\ 1 \leq j \leq d$, are $[0, 1]$ uniform random variables;

4. the $U_{i_1,...,i_d,j}$'s and η_k's are all stochastically independent.

Hence an unbiased estimator for μ based on a latin hypercube sampling is

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} f(X(\eta_1(k), \eta_2(k), ..., \eta_d(k))).$$

McKay, Beckman and Conover (McKay, M.D., 1979) futher proved that in a large number of instances, the variance of $\hat{\mu}_n$ is substantially smaller than that the estimators based on simple random sampling. Many years later, Stein (Stein, M.L., 1987) showed that the asymptotic variance of $\hat{\mu}_n$ is less than the asymptotic variance of an analogous estimator based on an independently and identically distributed sample. Later, Owen (Owen, A.B., 1992) gave the multivariate central limit theorem for $\hat{\mu}_n$ when f is bounded. In addition to the LHS, there are several ways to sample X_i's in order to estimate μ, namely, lattice sampling (Patterson, H.D., 1954), the orthogonal array sampling ((Loh, W.L., 1996a), (Neammanee, K. & Laipaporn, K., 2008), (Tang, B., 1993)), and scrambled net sampling ((Owen, A.B., 1997a), (Owen, A.B., 1997b)).

If $Var(\hat{\mu}_n) > 0$, we define

$$W = \frac{\hat{\mu}_n - \mu}{\sqrt{Var(\hat{\mu}_n)}}.$$

Then

$$EW = 0 \text{ and } VarW = 1. \tag{1}$$

To use the Stein's method to approximate the distribution of W with the standard normal distribution, Loh (Loh, W.L., 1996b) wrote

$$W = \sum_{i=1}^{n} V(\eta_1(i), \ldots, \eta_d(i)),$$

where

$$V(i_1, ..., i_d) = \frac{1}{n\sqrt{Var\hat{\mu}_n}}[f \circ X(i_1, ..., i_d) - \sum_{k=1}^{d} \mu_{-k}(i_k) + (d-1)\mu] \text{ and}$$

$$\mu(i_1, ..., i_d) = Ef \circ X(i_1, ..., i_d) \text{ and } \mu_{-k}(i_k) = \frac{1}{n^{(d-1)}} \sum_{j \neq k} \sum_{i_j=1}^{n} \mu(i_1, ..., i_d),$$

and gave the rate of convergence $\frac{C_d}{\sqrt{n}}$ without the value of C_d under the finiteness of the absolute third moment. Theorem 1.1 is his result.

Theorem 1.1 There exists a positive constant C_d which depends only on d such that for sufficiently large n,

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| \leq C_d \beta_3,$$

where Φ is the standard normal distribution and $\beta_3 = \frac{1}{n^{d-1}} \sum_{i_1=1}^{n} \ldots \sum_{i_d=1}^{n} E|V(i_1, ..., i_d)|^3$.

Corollary 1.2 If $E|f \circ X|^3 < \infty$, then

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| \leq \frac{C_d}{\sqrt{n}}.$$

In 2006, Rattanawong (Rattanawong, P., 2006) showed that there exist random permutations $\pi_1, \ldots, \pi_{d-1}$ on $\{1, 2, \ldots, n\}$ which are uniformly distributed over all the $n!$ possible permutations such that

$$W = \sum_{i=1}^{n} Y(i, \pi_1(i), \ldots, \pi_{d-1}(i))$$

and $Y(i_1, \ldots, i_d)$'s and π_k's are stochastically independent. Indeed, for $j \in \{1, \ldots, d\}$, let $\pi_j(\omega) = \eta_{j+1}(\omega)(\eta_1(\omega)^{-1})$ and for each $i_1, ..., i_d \in \{1, ..., n\}$, define

$$Y(i_1, ..., i_d) = \frac{1}{n\sqrt{var(\hat{\mu}_n)}}[f \circ X(i_1, ..., i_d) + \sum_{k=1}^{d-1} U_k(i_1, ..., i_d) + (-1)^d \mu], \tag{2}$$

where

$$\mu(i_1, ..., i_d) = Ef \circ X(i_1, ..., i_d), \quad U_k(i_1, ..., i_d) = \frac{(-1)^k}{n^k} \sum_{1 \leq j_1 < j_2 < ... < j_k \leq d} \sum_{q_{j_1}=1}^{n} \ldots \sum_{q_{j_k}=1}^{n} \mu(l_1, ..., l_d),$$

and

$$l_p = \begin{cases} q_p & \text{if } p = j_1, ..., j_k, \\ i_p & \text{otherwise.} \end{cases}$$

Note that the definition of $Y(i_1, ..., i_d)$'s are different from Loh (Loh, W.L., 1996b) in order that the random variable W satisfies the following property:

$$\sum_{i_j=1}^{n} EY(i_1, \ldots, i_d) = 0 \text{ for each } j \in \{1, 2, \ldots, n\}. \tag{3}$$

Furthermore, Neammanee and Rattanawong (Neammanee, K. & Rattanawong, P., 2008) used a concentration inequality approach of Stein's method and assumed the finiteness of fourth moment to give a constant C_d. This is their result.

Theorem 1.3 Suppose that $E(f \circ X(i_1, ..., i_d))^4 < \infty$, $1 \le i_1, ..., i_d \le n$. Then for $n \ge 6^d + 3$,

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le (11.765 + 23.531d)\delta_4 + \frac{11.68}{\sqrt{n}} + \frac{2.075d\delta_4^{\frac{1}{4}}}{n^{\frac{3}{8}}} + \frac{(2\sqrt{2\pi} + 10.027d)\delta_4^{\frac{3}{4}}}{n^{\frac{1}{8}}},$$

where Φ is the standard normal distribution and $\delta_4 = \frac{1}{n^{d+\frac{3}{2}}} \sum_{i_1=1}^n \cdots \sum_{i_d=1}^n E|Y(i_1, ..., i_d)|^4$.

Corollary 1.4 Suppose that $E(f \circ X(i_1, ..., i_d))^4 < \infty$, $1 \le i_1, ..., i_d \le n$. If $\delta_4 \sim \frac{1}{\sqrt{n}}$, then

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le \frac{28.729 + 35.633d}{\sqrt{n}}.$$

In this article, we use a concentation inequality approach of Stein's method with ideas of Neammanee and Rattanawong ((Neammanee, K. & Rattanawong, P., 2008), (Neammanee, K. & Rattanawong, P., 2009b)) and Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.) to obtain a constant C_d by assuming the finiteness of the absolute third moment. Theorem 1.5 is our main result.

Theorem 1.5 Suppose that $E|f \circ X(i_1, ..., i_d)|^3 < \infty$, $1 \le i_1, ..., i_d \le n$. For $n \ge 6^d$,

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le (22.88 + 28.99d)\delta_3 + \frac{3.88 + 2.09d}{\sqrt{n}} + 1.03\delta_3^2 + (\frac{C_d}{n^{\frac{1}{24}}})(\frac{C\delta_3^{\frac{3}{4}}}{n^{\frac{1}{8}}} + Cn^{\frac{1}{8}}\delta_3^{\frac{5}{4}}) + O(\frac{1}{n}),$$

where

$$\delta_3 = \frac{1}{n^{d-1}} \sum_{i_1=1}^n \cdots \sum_{i_d=1}^n E|Y(i_1, ..., i_d)|^3$$

and the definition of $Y(i_1, ..., i_d)$ is given by (2).

Corollary 1.6 Suppose that $E|f \circ X(i_1, ..., i_d)|^3 < \infty$, $1 \le i_1, ..., i_d \le n$. If $n \ge 6^d$ and $\delta_3 \sim \frac{1}{\sqrt{n}}$, then

$$\sup_{z\in\mathbb{R}} |P(W \le z) - \Phi(z)| \le \frac{26.76 + 31.08d}{\sqrt{n}} + \frac{2.92d}{n} + O(\frac{1}{n^{\frac{13}{24}}}).$$

Example. In the case of $d = 2$, we observe that this is a special case of the combinatorial central limit theorem (For more detail see Von Bahr (Von Bahr, B., 1976), Ho and Chen (Ho, S.T., 1978)). Under the finiteness of absolute third moment, Neammanee and Suntornchost (Neammanee, K. & Suntornchost, J., 2005) gave the uniform rate of convergence and obtained the rate $\frac{198}{\sqrt{n}}$. Recently, Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.) improve the constant to be 78.36. For this work, Corollary 1.6 yields the constant 93.17. Althought this constant is not shaper than the previous result, we establish a uniform bound on a generalization of a combinatorial central limit theorem by assuming the finiteness of absolute third moment.

2. Auxiliary Results

In this section, we will give some lemmas which are used in the next section. Almost of them, we generalize the results of Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.) and improve the results of Neammanee and Rattanawong (Neammanee, K. & Rattanawong, P., 2009b) under the finiteness of absolute third moment. Throughtout this work, we let

$$\delta_2 = \frac{1}{n^{d-\frac{1}{2}}} \sum_{i_1=1}^n \cdots \sum_{i_d=1}^n E|Y(i_1, ..., i_d)|^2 \text{ and } \delta_3 = \frac{1}{n^{d-1}} \sum_{i_1=1}^n \cdots \sum_{i_d=1}^n E|Y(i_1, ..., i_d)|^3.$$

Lemma 2.1 Suppose that $E|f \circ X(i_1, ..., i_d)|^3 < \infty$, $1 \le i_1, ..., i_d \le n$. For $n \ge 36$,

$$\delta_2 \le \frac{1.02943}{\sqrt{n}}.$$

Proof. By (1), we have

$$
\begin{aligned}
1 &= EW^2 \\
&= \sum_{i=1}^{n} EY^2(i,\pi_1(i),\ldots,\pi_{d-1}(i)) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} EY(i,\pi_1(i),\ldots,\pi_{d-1}(i))Y(j,\pi_1(j),\ldots,\pi_{d-1}(j)) \\
&= \sqrt{n}\delta_2 + \frac{1}{n^{d-1}(n-1)^{d-1}} \sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}\sum_{\substack{j_1=1\\ j_1\neq i_1}}^{n}\cdots\sum_{\substack{j_d=1\\ j_d\neq i_d}}^{n} EY(i_1,\ldots,i_d)EY(j_1,\ldots,j_d) \\
&= \sqrt{n}\delta_2 + \frac{(-1)^d}{n^{d-1}(n-1)^{d-1}} \sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}[EY(i_1,\ldots,i_d)]^2.
\end{aligned}
$$

Thus

$$
\sqrt{n}\delta_2 \leq 1 + \frac{1}{n^{d-1}(n-1)^{d-1}} \sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}[EY(i_1,\ldots,i_d)]^2 \leq 1 + \frac{\sqrt{n}\delta_2}{(n-1)^{d-1}}.
$$

This implies that for $n \geq 36$ and $d \geq 2$,

$$
\sqrt{n}\delta_2 \leq 1 + \frac{\sqrt{n}\delta_2}{35^2} = 1 + 0.02858\,\sqrt{n}\delta_2
$$

and hence

$$
\delta_2 \leq \frac{1}{(1-0.02858)\sqrt{n}} \leq \frac{1.02943}{\sqrt{n}}.
$$

□

For each $i_1,\ldots,i_d \in \{1,2,\ldots,n\}$, we define

$$
Y_0(i_1,\ldots,i_d) = Y(i_1,\ldots,i_d)\mathbb{I}(|Y(i_1,\ldots,i_d)|>1),\ \widehat{Y}_0(i_1,\ldots,i_d) = Y(i_1,\ldots,i_d)\mathbb{I}(|Y(i_1,\ldots,i_d)|\leq 1),\quad \text{and}
$$

$$
\widehat{Y}(\pi) = \sum_{i=1}^{n} \widehat{Y}_0(i,\pi_1(i),\ldots,\pi_{d-1}(i)),
$$

where $\mathbb{I}$ is the indicator function, i.e., for a nonempty set A, the indicator function of A is defined by

$$
\mathbb{I}(A)(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}
$$

Next, we note that for any integer m, n and r which $m \geq 0$, and $n, r > 0$,

$$
\begin{aligned}
E|Y^m(i_1,\ldots,i_d)Y_0^n(i_1,\ldots,i_d)| &\leq E|Y^m(i_1,\ldots,i_d)Y_0^n(i_1,\ldots,i_d)||Y_0(i_1,\ldots,i_d)|^r \\
&\leq E|Y(i_1,\ldots,i_d)|^{m+n+r}.
\end{aligned} \tag{4}
$$

Lemma 2.2 Suppose that $E|f\circ X(i_1,...,i_d)|^3 < \infty$, $1 \leq i_1,...,i_d \leq n$. If $n \geq 36$, then

$$
E[\sum_{i=1}^{n}\sum_{k=1}^{n} Y(i,\pi_1(k),\ldots,\pi_{d-1}(k))]^2 \leq 1.02943n.
$$

Proof. Observe that

$$
\begin{aligned}
&E[\sum_{i=1}^{n}\sum_{k=1}^{n} Y(i,\pi_1(k),\ldots,\pi_{d-1}(k))]^2 \\
&\quad= \sum_{i=1}^{n}\sum_{k=1}^{n} EY^2(i,\pi_1(k),\ldots,\pi_{d-1}(k)) + \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{\substack{l=0\\ (i,k)\neq(l,m)}}^{n}\sum_{m=0}^{n} EY(i,\pi_1(k),\ldots,\pi_{d-1}(k))Y(l,\pi_1(m),\ldots,\pi_{d-1}(m))
\end{aligned}
$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{n} EY^2(i,\pi_1(k),\ldots,\pi_{d-1}(k)) + \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{\substack{m=1\\ m\neq k}}^{n} EY(i,\pi_1(k),\ldots,\pi_{d-1}(k))Y(l,\pi_1(m),\ldots,\pi_{d-1}(m))$$

$$+ \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n} EY(i,\pi_1(k),\ldots,\pi_{d-1}(k))Y(l,\pi_1(k),\ldots,\pi_{d-1}(k))$$

$$= \frac{1}{n^{d-2}}\sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n} EY^2(i_1,\ldots,i_d) + \frac{1}{(n(n-1))^{d-2}}\sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}\sum_{\substack{l_2=1\\ l_2\neq i_2}}^{n}\cdots\sum_{\substack{l_d=1\\ l_d\neq i_d}}^{n} EY(i_1,\ldots,i_d)\sum_{l_1=1}^{n} EY(l_1,\ldots,l_d)$$

$$+ \frac{1}{n^{d-2}}\sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}\sum_{\substack{l=1\\ l\neq i_1}}^{n} EY(i_1,\ldots,i_d)EY(l,i_2,\ldots,i_d).$$

By Lemma 2.1, (3) and (4), we have

$$E[\sum_{i=1}^{n}\sum_{k=1}^{n} Y(i,\pi_1(k),\ldots,\pi_{d-1}(k))]^2 \leq 1.02943n.$$

□

From now, we use the following system giving by Ho and Chen (Ho, S.T., 1978) and Neammanee and Rattanawong (Neammanee, K. & Rattanawong, P., 2008). Let $I, K, L_1,\ldots,L_{d-1}, M_1,\ldots, M_{d-1}$ be uniformly distributed random variables on $\{1,2,\ldots,n\}$ and $\rho_1,\ldots,\rho_{d-1}$ and $\tau_1,\ldots,\tau_{d-1}$ are random permutations of $\{1,2,\ldots,n\}$. Assume that

$\{I,K,L_1,\ldots,L_{d-1},M_1,\ldots,M_{d-1},\rho_1,\ldots,\rho_{d-1},\tau_1,\ldots,\tau_{d-1}\}$ is independent of $Y(i_1,i_2)$'s,

$(I,K),(L_1,M_1),\ldots,(L_{d-1},M_{d-1})$ are uniformly distributed on $\{(i,k)|i,k=1,2,...,n$ and $i\neq k\}$,

$(I,K),(L_1,M_1),\ldots,(L_{d-1},M_{d-1})$ and $\tau_1,\ldots,\tau_{d-1}$ are mutually independent,

(I,K) and $\rho_1,\ldots,\rho_{d-1}$ are mutually independent, and

$$\rho_i(\alpha) = \begin{cases} \tau_i(\alpha) & \text{if } \alpha\neq I,K,\tau_i^{-1}(L),\tau_i^{-1}(M),\\ L_i & \text{if } \alpha = I,\\ M_i & \text{if } \alpha = K,\\ \tau_i(I) & \text{if } \alpha = \tau_i^{-1}(L),\\ \tau_i(K) & \text{if } \alpha = \tau_i^{-1}(M),\end{cases}$$

where $\rho_i(\rho_i^{-1}(\alpha)) = \rho_i^{-1}(\rho_i(\alpha)) = \alpha$ for $i = 1,2,\ldots,d-1$. Now, we define some notations;

$$\widetilde{Y}(\rho) = \widehat{Y}(\rho) - \widehat{S}_{1,0} - \widehat{S}_{2,0} + \widehat{S}_{3,0} + \widehat{S}_{4,0},$$

where

$$\widehat{Y}(\rho) = \sum_{i=1}^{n}\widehat{Y}_0(i,\rho_1(i),\ldots,\rho_{d-1}(i)),$$

$$\widehat{S}_{1,0} = \widehat{Y}_0(I,\rho_1(I),\ldots,\rho_{d-1}(I)),\quad \widehat{S}_{2,0} = \widehat{Y}_0(K,\rho_1(K),\ldots,\rho_{d-1}(K)),$$

$$\widehat{S}_{3,0} = \widehat{Y}_0(I,\rho_1(K),\ldots,\rho_{d-1}(K)),\quad \widehat{S}_{4,0} = \widehat{Y}_0(K,\rho_1(I),\ldots,\rho_{d-1}(I)).$$

It is easy to see that $\widehat{S}_{1,0},\widehat{S}_{2,0},\widehat{S}_{3,0},\widehat{S}_{4,0}$ have the same distribution.

Lemma 2.3 Suppose that $E|f\circ X(i_1,...,i_d)|^3<\infty$, $1\leq i_1,...,i_d\leq n$. Then

(1) $E\widehat{Y}^2(\pi) \leq 1 + \dfrac{2\sqrt{n}\delta_2}{(n-1)^{d-1}} + \dfrac{n^{d-1}\delta_3^2}{(n-1)^{d-1}}$.

(2) For $n\geq 6^d$ and $\delta_3\leq \dfrac{1}{30}$, we have $E\widehat{Y}^4(\pi)\leq C(n+n^2\delta_3^2)$.

(3) $E|\widetilde{Y}(\rho)-\widehat{Y}(\rho)|^2 = \dfrac{4}{n} + \mathcal{R}$, where

$$|\mathcal{R}| \leq \frac{36\delta_2}{\sqrt{n}(n-1)} + \frac{8\delta_3}{n-1} + \frac{4n^{d-1}\delta_2^2}{(n-1)^d} + \frac{4n^{d-2}\delta_3^2}{(n-1)^{d-1}}.$$

Proof. By using the same argument of Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.), we have this lemma. □

Corollary 2.4 Suppose that $E|f \circ X(i_1, ..., i_d)|^3 < \infty$, $1 \le i_1, ..., i_d \le n$. For $n \ge 6^d$ and $\delta_3 \le \frac{1}{30}$, we get that

(1) $E\widehat{Y}^2(\pi) \le 1.05998$.

(2) $E|\widetilde{Y}(\rho) - \widehat{Y}(\rho)|^2 \le \frac{4.27125}{n-1} + O(\frac{1}{n^2})$.

Proof. We apply the idea of Neammanee and Rattanawong (Neammanee, K. & Rattanawong, P., 2008) and the fact that $\frac{d-1}{n-1} \le \frac{1}{35}$ for all $n \ge 6^d$, we have

$$(\frac{n}{n-1})^{d-1} \le 1 + \sum_{r=1}^{\infty}(\frac{d-1}{n-1})^r \le 1.03. \tag{5}$$

By Lemma 2.1 and (5), we obtain this corollary. □

3. Proof of Theorem 1.5

We will prove this theorem by using ideas in two papers of Neammanee and Rattanawong ((Neammanee, K. & Rattanawong, P., 2008), (Neammanee, K. & Rattanawong, P., 2009b)) and a paper of Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.). Since $|P(W \le z) - \Phi(z)| \le 0.55$ (Chen, L.H.Y., 2001, p. 246), we can assume $\delta_3 \le \frac{1}{30}$. Assume that $z > 0$. In case of $z < 0$, we use the fact that $\Phi(z) = 1 - \Phi(-z)$ and then apply the result to $-W$. Using the same argument of Neammanee and Rattanawong (Neammanee, K. & Rattanawong, P., 2009b), we have

$$\begin{aligned}|P(W \le z) - \Phi(z)| &\le P(W \ne \widehat{Y}(\pi)) + |P(\widehat{Y}(\pi) \le z) - \Phi(z)| \\ &\le \delta_3 + |T_1| + |T_2| + |T_3| + |T_4|,\end{aligned}$$

where

$$\begin{aligned}T_1 &= Eg_z'(\widehat{Y}(\tau))\int_{-\infty}^{\infty}K(t)dt - E\int_{-\infty}^{\infty}g_z'(\widehat{Y}(\rho)+t)K(t)dt, \\ T_2 &= Eg_z'(\widehat{Y}(\tau))E\int_{-\infty}^{\infty}K(t)dt - Eg_z'(\widehat{Y}(\tau))\int_{-\infty}^{\infty}K(t)dt, \\ T_3 &= Eg_z'(\widehat{Y}(\tau)) - Eg_z'(\widehat{Y}(\tau))E\int_{-\infty}^{\infty}K(t)dt, \\ T_4 &= \frac{1}{n}Eg_z(\widehat{Y}(\rho))\sum_{i=1}^{n}\sum_{k=1}^{n}\widehat{Y}_0(i, \rho_1(k), \dots, \rho_{d-1}(k)), \\ K(t) &= \frac{n-1}{4}(\widetilde{Y}(\rho) - \widehat{Y}(\rho))(\mathbb{I}(0 \le t \le \widetilde{Y}(\rho) - \widehat{Y}(\rho)) - \mathbb{I}(\widetilde{Y}(\rho) - \widehat{Y}(\rho) \le t < 0))\end{aligned}$$

and g_z is the solution of the Stein's equation for normal distribution function

$$g'(w) - wg(w) = \mathbb{I}(w \le z) - \Phi(z), \quad \text{for all } w \in \mathbb{R}.$$

First, we bound T_4 by using Lemma 2.2 and the fact that $0 \le g_z(w)) \le \min(\frac{\sqrt{2\pi}}{4}, \frac{1}{|z|})$ (Chen, L.H.Y., 2001, p. 246). Indeed,

$$\begin{aligned}|T_4| &\le \frac{1}{n}E|g_z(\widehat{Y}(\rho))||\sum_{i=1}^{n}\sum_{k=1}^{n}Y(i, \rho_1(k), \dots, \rho_{d-1}(k))| + \frac{1}{n}E|g_z(\widehat{Y}(\rho))||\sum_{i=1}^{n}\sum_{k=1}^{n}Y_0(i, \rho_1(k), \dots, \rho_{d-1}(k))| \\ &\le \frac{1}{n}\{Eg_z^2(\widehat{Y}(\rho))\}^{\frac{1}{2}}\{E[\sum_{i=1}^{n}\sum_{k=1}^{n}Y(i, \rho_1(k), \dots, \rho_{d-1}(k))]^2\}^{\frac{1}{2}} + \frac{\sqrt{2\pi}}{4n}\sum_{i=1}^{n}\sum_{k=1}^{n}E|Y_0(i, \rho_1(k), \dots, \rho_{d-1}(k))| \\ &\le \frac{\sqrt{2\pi}}{4n}[\sqrt{1.02943n} + \frac{1}{n^{d-2}}\sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}E|Y(i_1, \dots, i_d)|^3] \\ &\le \frac{0.63582}{\sqrt{n}} + 0.62666\delta_3.\end{aligned}$$

We apply Lemma 2.3(3) and the fact that $|g'_z(w)| \leq 1$ (Stein, C.M., 1986, p. 23) to obtain

$$|T_3| \leq E|g'_z(\widehat{Y}(\tau))||1 - E\int_{-\infty}^{\infty} K(t)dt| \leq |1 - \frac{(n-1)E|\widetilde{Y}(\rho) - \widehat{Y}(\rho)|^2}{4}| \leq 2\delta_3 + 1.03\delta_3^2 + O(\frac{1}{n}).$$

Next, we will bound T_2. Let $\mathcal{B}$ be the σ-algebra generated by

$$\{I, K, L_1, \ldots, L_{d-1}, M_1, \ldots, M_{d-1}, Y(i_1, \ldots, i_d) : 1 \leq i_1, i_2, \ldots, i_d \leq n\}$$
$$A = \{\tau_i(I) \neq L_i, \tau_i(K) \neq M_i, \tau_i(I) \neq M_i, \tau_i(K) \neq L_i, i = 1, \ldots, d-1\}, \text{ and}$$
$$\widehat{G} = \widehat{Y}_0(I, M_1, \ldots, M_{d-1}) + \widehat{Y}_0(K, L_1, \ldots, L_{d-1}) - \widehat{Y}_0(I, L_1, \ldots, L_{d-1}) - \widehat{Y}_0(K, M_1, \ldots, M_{d-1}).$$

Observe that this is a generalization of the definition of Neamanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.). By the same argument of Neammanee and Rattanawong (Neammanee, K. & Rattanawong, P., 2008, p. 24-25), we have

$$E^{\mathcal{B}}\mathbb{I}(A^c) \leq \frac{4}{n}\sum_{r=1}^{n}\binom{d-1}{r} \leq \frac{4}{n}(2^{d-1}-1) \text{ and } |T_2| \leq \frac{n-1}{2}E[\widehat{G}^2 E^{\mathcal{B}}\mathbb{I}(A^c)].$$

By Lemma 2.3(3), we get that

$$|T_2| \leq \frac{2(n-1)}{n}\sum_{r=1}^{n}\binom{d-1}{r}E|\widetilde{Y}(\rho) - \widehat{Y}(\rho)|^2 = O(\frac{1}{n}).$$

Finally, we will bound T_1. Denote $\Delta\widehat{Y} = \widehat{Y}(\rho) - \widehat{Y}(\tau)$. Again, by using the same argument of Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.), we have

$$T_1 \leq B_1 + B_2 + B_3 + B_4 + B_5, \tag{6}$$

where

$$B_1 = E\int_{\substack{\widehat{Y}(\tau)<z \\ \widehat{Y}(\tau)+\Delta\widehat{Y}+t>z}} K(t)dt, \quad B_2 = E\int_{\mathbb{R}} |\widehat{Y}(\tau)||\Delta\widehat{Y}|K(t)dt, \quad B_3 = E\int_{\mathbb{R}} |\widehat{Y}(\tau)||t|K(t)dt,$$
$$B_4 = \frac{\sqrt{2\pi}}{4}E\int_{\mathbb{R}} |\Delta\widehat{Y}|K(t)dt, \quad B_5 = \frac{\sqrt{2\pi}}{4}E\int_{\mathbb{R}} |t|K(t)dt, \text{ and}$$
$$|B_2 + B_3 + B_4 + B_5| \leq 26.9862d\delta_3 + 13.24969\delta_3 + (\frac{C_d}{n^{\frac{1}{24}}})[\frac{C\delta_3^{\frac{3}{4}}}{n^{\frac{1}{8}}} + Cn^{\frac{1}{8}}\delta_3^{\frac{5}{4}}] + O(\frac{1}{n}). \tag{7}$$

Hence, it suffices to bound B_1. For each $\delta \geq 0$ and $a, b \in \mathbb{R}$, where $a < b$, we define the function f_δ by

$$f_\delta(t) = \begin{cases} -\frac{1}{2}(b-a) - \delta & \text{if } t < a - \delta, \\ -\frac{1}{2}(b+a) + t & \text{if } a - \delta \leq t \leq b + \delta, \\ \frac{1}{2}(b-a) + \delta & \text{if } b + \delta < t. \end{cases}$$

Then

$$|f_\delta(t)| \leq \frac{1}{2}(b-a) + \delta \text{ for every } t \in \mathbb{R}. \tag{8}$$

Note that $E|\Delta\widehat{Y}|^k$ is bounded by a sum of $(4d)^k$ terms each of the form $E|\widehat{Y}_0(I, \rho_1(I), \ldots, \rho_{d-1}(I))|^k$. This implies

$$E|\Delta\widehat{Y}| \leq 4d \quad \text{and} \quad E|\Delta\widehat{Y}|^k \leq \frac{(4d)^k\delta_k}{n^{\frac{k-1}{2}}} \tag{9}$$

for $k \in \{2, 3\}$. Similar to T_4, we obtain

$$|\Delta f_\delta(\widehat{Y}(\rho))| \leq \frac{1.01461}{\sqrt{n}}(Ef_\delta^2(\widehat{Y}(\rho)))^{\frac{1}{2}} + (2d+6)\delta_3. \tag{10}$$

By the same argument of Neammanee and Rerkruthairat (Neammanee, K. & Rerkruthairat, N.) Lemma 2.1, 2.3(1, 3) and (8) to (10), we have

$$B_1 \leq \frac{3.23695}{\sqrt{n}} + \frac{2.08919d}{\sqrt{n}} + (2d+6)\delta_3 + O(\frac{1}{n}). \tag{11}$$

We use (6), (7) and (11) to conclude that

$$T_1 \leq \frac{3.23695}{\sqrt{n}} + \frac{2.08919d}{\sqrt{n}} + 28.9862d\delta_3 + 19.24969\delta_3 + (\frac{C_d}{n^{\frac{1}{24}}})[\frac{C\delta_3^{\frac{3}{4}}}{n^{\frac{1}{8}}} + Cn^{\frac{1}{8}}\delta_3^{\frac{5}{4}}] + O(\frac{1}{n}). \tag{12}$$

By the same argument of (12), we have

$$T_1 \geq \frac{-3.23695}{\sqrt{n}} - \frac{2.08919d}{\sqrt{n}} - 28.9862d\delta_3 - 19.24969\delta_3 - (\frac{C_d}{n^{\frac{1}{24}}})[\frac{C\delta_3^{\frac{3}{4}}}{n^{\frac{1}{8}}} + Cn^{\frac{1}{8}}\delta_3^{\frac{5}{4}}] - O(\frac{1}{n})$$

(see (Neammanee, K. & Rattanawong, P., 2008, p. 13) for more detail). Hence

$$|T_1| \leq \frac{3.23695}{\sqrt{n}} + \frac{2.08919d}{\sqrt{n}} + 28.9862d\delta_3 + 19.24969\delta_3 + (\frac{C_d}{n^{\frac{1}{24}}})[\frac{C\delta_3^{\frac{3}{4}}}{n^{\frac{1}{8}}} + Cn^{\frac{1}{8}}\delta_3^{\frac{5}{4}}] + O(\frac{1}{n}).$$

Therefore,

$$|P(W \leq z) - \Phi(z)| \leq 22.87635\delta_3 + \frac{3.87277}{\sqrt{n}} + \frac{2.08919d}{\sqrt{n}} + 28.9862d\delta_3 + 1.03\delta_3^2 + (\frac{C_d}{n^{\frac{1}{24}}})[\frac{C\delta_3^{\frac{3}{4}}}{n^{\frac{1}{8}}} + Cn^{\frac{1}{8}}\delta_3^{\frac{5}{4}}] + O(\frac{1}{n}).$$

Acknowledgment

The authors would like to thank the Commission on Higher Education for financial support.

References

Chen, L.H.Y. & Shao, Q.M. (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Relat. Fields.*, 120, 236-254.

Ho, S.T. & Chen, L.H.Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.*, 6, 231-249.

Loh, W.L. (1996a). A combinatorial central limit theorem for randomized orthogonal array sampling designs. *Ann. Statist.*, 24, 1209-1224.

Loh, W.L. (1996b). On latin hypercube sampling. *Ann. Statist.*, 24, 2058-2080.

McKay, M.D., Conover, W.J. & Beckman, R.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21, 239-245.

Neammanee, K. & Laipaporn, K. (2008). A uniform bound on a combinatorial central limit theorem for randomized orthogonal array sampling designs. *Stoch. Anal. Appl.*, 26, 243-255.

Neammanee, K. & Rattanawong, P. (2008). A uniform bound on the generalization of a combinatorial central limit theorem. *Int. Math. Forum.*, 3, 11-27.

Neammanee, K. & Rattanawong, P. (2009a). A constant on a uniform bound of a combinatorial central limit theorem. *JMR*, 1, 91-103.

Neammanee, K. & Rattanawong, P. (2009b). Non-uniform bound on normal approximation of latin hypercube sampling. *JMR*, 1, 28-42.

Neammanee, K. & Rerkruthairat, N. An improvement of a uniform bound on a combinatorial central limit theorem. (to be appear in Communications in Statistics-Theory and Methods).

Neammanee, K. & Suntornchost, J. (2005). A uniform bound on a combinatorial central limit theorem. *Stoch. Anal. Appl.*, 23, 559-578.

Owen, A.B. (1992). A central limit theorem for latin hypercube sampling. *J.R. Statist. Soc. Ser. B*, 54, 541-551.

Owen, A.B. (1997a). Monte-Carlo variance of scrambled net quadrature. *SIAM J. Numer.Anal.*, 34, 1884-1910.

Owen, A.B. (1997b). Scrambled net variance for integrals of smooth functions. *Ann. Statist.*, 25, 1541-1562.

Patterson, H.D. (1954). The errors of lattice sampling. *J.R. Statist. Soc. Ser. B*, 16, 140-149.

Rattanawong, P. (2006). *Bounds on a normal approximation for latin hypercube sampling*, Ph.D. thesis, Chulalongkorn University.

Stein, C.M. (1986). *Approximate Computation of Expectation*. California:MIS, Hayward.

Stein, M.L. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29, 143-151.

Tang, B. (1993). Orthogonal array-based latin hypercubes. *J. Amer. Statist. Assoc.*, 88, 1392-1397.

Von Bahr, B. (1976). Remainder term estimate in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete.*, 35, 131-139.