U-type and factorial designs for nonparametric Bayesian regression

Rong-Xian Yue, Jing-Wen Wu
Department of Applied Mathematics, Shanghai Normal University,
100 Guilin Road, Shanghai 200234, China

Abstract

This paper deals with the design problem for recovering a response surface by using a nonparametric Bayesian approach. The criterion for selecting the designs is based on the asymptotic average estimation variance, and three priors for the response are specified. We found the optimal design that minimizes the criterion over the lattice designs with s q-level factors and N runs. The approach we used is similar to that in Ma et al. (2003). We also obtained alternative expressions and lower bounds for the criterion corresponding to each of the three Bayes models for the two-level U-type design by using the column balance and row distance proposed in Fang et al. (2003).

AMS subject classification: 62K05, 62J02.

Key words: Bayesian model, factorial experiment, response surface design.

1. Introduction

This paper deals with the design problem of recovering a response surface $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X} \subset \mathcal{R}^s$, from observations of f on a discrete set, $\xi = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, of points in $\mathcal{X}$ (called the design). The treatment is Bayesian in that knowledge of the response is represented by a random function.

Traditional Bayesian design theory deals with the design problem for fitting a linear regression model, where the prior of the response is a finite linear combination of known functions. Dealing with the infinite dimensional problems, Steinberg(1985) investigated the problem of a model-robust design for response surface study. In that paper, a Bayesian model is proposed

[*] This work was partially supported by a NSFC grant 10271078, and the Special Funds for Major Specialties of Shanghai Education Committee.

that makes explicit assumptions about the inadequacy of an assumed model. The specification on the prior distribution is based on the tensor products of Hermite polynomials. The design criterion based on the model leads to reasonable choices of scale for two-level factorial designs. Mitchell et al. (1994) studied the Bayesian model in which the prior may be infinite dimensional. Specifically, they assume that the prior is the one which derives from taking the response to have the form $y(\boldsymbol{x}) = \mu + Z(\boldsymbol{x}) + \varepsilon$, where μ is a normal random variable, Z is a Gaussian process independent of μ, and ε is a random error. They give three criteria, which are called the D, G and A criteria, by using the asymptotic analysis from allowing the error variance to be large. Yue (2001) described a Bayesian model in which the prior for the response is specified based on a functional ANOVA decomposition. A criterion is developed there by using the mechanics for asymptotic used in Mitchell et al. (1994), and a comparison of some random and quasi-random point sets is given the Bayesian design.

In this paper, we still use the mechanics for asymptotic in Mitchell et al. (1994), and consider U-type designs for the nonparametric Bayesian model. A q-level U-type design $U(N; q^s)$ is an $N \times s$ matrix, with each column having equal number of $\frac{2\ell-1}{2q}$, $\ell = 1, \cdots, q$. By a linear transformation a $\frac{2\ell-1}{2q} \to \ell$, a $U(N; q^s)$ can also be presented as a matrix of size $N \times s$, with each column having equal number of $\ell = 1, \cdots, q$. Let $\mathcal{U}(N; q^s)$ be the set of $U(N; q^s)$'s. Most of the uniform design are constructed based on U-type designs. Ma et al. (2003) obtained exact conditions in which the uniform U-type design can imply design orthogonality. Fang et al. (2003) studied the uniformity of two- and three-level U-type designs based on the centered and wrap-around L_2-discrepancies. They developed some new representations and lower bounds of the L_2-discrepancies in terms of column balance and Hamming distances of the rows, respectively.

In the spirit of formulation presented by Mitchell et al. (1994), we take five covariance kernels into account for the response. For each kernel, the asymptotic expected estimation variance is reexpressed as functions of column balance, and also as functions of Hamming distances of the rows by using the approach of Fang et al. (2003). And the lower bounds on the asymptotic expected estimation variance are given, which can be used as bench marks in searching U-type design for the Bayesian model.

In Section 2, we describe the Bayesian models and the criterion for choosing designs based on the asymptotic expected estimation variance. In Section 3, we found the optimal design that minimizes the criterion over the lattice designs with s q-level factors and N runs. In Section 4, We derive alternative expressions and lower bounds for the criterion corresponding to each of the three Bayes models for the two-level U-type design by using the column balance and row distance. A Summary is given in Section 5.

2. Asymptotic Bayes criterion

According to Mitchell et al. (1994), we assume that the prior is the one which derives from taking the response have the form

$$f(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}), \tag{1}$$

where β is a normal random variable with mean 0 and variance σ_β^2, and Z is a Gaussian process independent of β, with mean 0 and covariance function

$$\mathrm{cov}[Z(\boldsymbol{x}), Z(\boldsymbol{t})] = \eta K(\boldsymbol{x}, \boldsymbol{t}),$$

where $\eta \in (0, \infty)$ is a parameter.

Let y_i represent an observation taken at the experimental setting $\boldsymbol{x}_i \in \mathcal{X}$,

$$y_i = f(\boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, \cdots, N,$$

where the terms ε_i $(i = 1, \cdots, N)$ are uncorrelated random variables with mean 0 and constant variance σ_ε^2. Let $\boldsymbol{y} = (y_1, \ldots, y_N)'$. Then the prior specification above implies that $f(\boldsymbol{x})$ will have a normal posterior distribution. Thus we estimate f by its posterior expectation, $E[f(\boldsymbol{x})|\boldsymbol{y}]$, and assess the estimation accuracy by its posterior variance, $\mathrm{var}[f(\boldsymbol{x})|\boldsymbol{y}]$, which we call estimation variance.

Let $\boldsymbol{k}(\boldsymbol{x})$ be the N-vector whose i-th component is $K(\boldsymbol{x}, \boldsymbol{x}_i)$, i.e.,

$$\boldsymbol{k}(\boldsymbol{x}) = (K(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_N))',$$

and let $\boldsymbol{K}$ be the $(N \times N)$-matrix whose (i, j)-th element is $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Then the estimation variance for each $\boldsymbol{x} \in \mathcal{X}$ is

$$\mathrm{var}[f(\boldsymbol{x})|\boldsymbol{y}] = \eta \left[\nu + K(\boldsymbol{x}, \boldsymbol{x}) - (\nu \boldsymbol{1}_N' + \boldsymbol{k}(\boldsymbol{x})')(\nu \boldsymbol{1}_N \boldsymbol{1}_N' + \boldsymbol{K} + \omega \boldsymbol{I}_N)^{-1}(\nu \boldsymbol{1}_N + \boldsymbol{k}(\boldsymbol{x}))\right],$$

where $\nu = \sigma_\beta^2/\eta$, $\omega = \sigma_\varepsilon^2/\eta$, and $\boldsymbol{1}_N$ is the N-vector of 1's. Note that ν and ω can be considered as the variance of β and ε_i relative to the variance of the random process Z, respectively.

A natural criterion by which to compare the estimation variance functions for different designs is a weighted average using some weighted function, $\rho(\boldsymbol{x})$. We define the average weighted estimation variance, V_{avg}, for an experimental design by

$$V_{\mathrm{avg}} = \sigma_\varepsilon^{-2} \int_{\mathcal{X}} \mathrm{var}[f(\boldsymbol{x})|\boldsymbol{y}] \rho(d\boldsymbol{x}).$$

Using the asymptotics in Mitchell et al. (1994), we allow both ν and ω to be large while $\gamma = \nu/\omega$ is held fixed. Defining

$$\alpha = 1/\omega, \quad \hat{\gamma} = \gamma/(1 + N\gamma),$$

the average estimation variance can then be expressed by

$$V_{\text{avg}} = \hat{\gamma} + \alpha \left[\int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}) \rho(d\boldsymbol{x}) - 2\hat{\gamma} \sum_{i=1}^{N} \int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}_i) \rho(d\boldsymbol{x}) + \hat{\gamma}^2 \sum_{i,j=1}^{N} K(\boldsymbol{x}_i, \boldsymbol{x}_j) \right] + O(\alpha^2).$$

Ignoring terms of order $O(\alpha^2)$, a good design $\xi = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ should be selected so that the following quantity, $\Psi(\xi; K)$, is as small as possible:

$$\Psi(\xi; K) = \int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}) \rho(d\boldsymbol{x}) - 2\hat{\gamma} \sum_{i=1}^{N} \int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}_i) \rho(d\boldsymbol{x}) + \hat{\gamma}^2 \sum_{i,j=1}^{N} K(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{2}$$

Note that if $\hat{\gamma} = 1/n$ then $\Psi(\xi; K)$ becomes a discrepancy of the point set ξ based on the reproducing kernel K except a constant term (Hickernell, 1998).

In what follows, we will assume that the experimental domain is the s-dimensional unit cube $[0,1]^s$ with uniform measure $\rho(d\boldsymbol{x}) = d\boldsymbol{x}$, and consider the following two covariance kernels:

$$\begin{aligned} K_1(\boldsymbol{x}, \boldsymbol{t}) &= \prod_{r=1}^{s} [\min(x_r, t_r) - x_r t_r], \\ K_2(\boldsymbol{x}, \boldsymbol{t}) &= \prod_{r=1}^{s} (1 - |x_r - t_r|), \\ K_3(\boldsymbol{x}, \boldsymbol{t}) &= \prod_{r=1}^{s} \left[1 + \theta \left(\frac{1}{3} + \frac{x_r^2 + t_r^2}{2} - \max(x_r, t_r) \right) \right], \end{aligned} \tag{3}$$

where $\theta \in (0,1)$ is a parameter that specifies the rate at which higher-order interactions are discounted.

To understand K_1 and K_2, we consider the one-dimensional case. For $s = 1$, $K_1(x,t) = \min(x,t) - xt$ is the covariance kernel of the Brownian bridge on the function class $\{h \in C([0,1]) :\ h(0) = h(1) = 0\}$. The covariance kernel K_2 is the covariance kernel of the sum $h_1(x) + h_2(1-x)$ where h_1 and h_2 are independent and distributed according to the Wiener measure on the class $\{h \in C([0,1]) :\ h(0) = 0\}$.

As to the kernel K_3, we consider a random function, $F(\boldsymbol{x})$, say. We express $F(\boldsymbol{x})$ in terms of its ANOVA decomposition (Owen, 1992) as follows:

$$F(\boldsymbol{x}) = \sum_{u \subseteq \{1,\ldots,s\}} F_u(\boldsymbol{x}_u),$$

where each $F_u(\boldsymbol{X}_u)$ is defined recurrently by

$$\begin{aligned} F_\emptyset &= \int_{[0,1]^s} F(\boldsymbol{x}) d\boldsymbol{x}, \\ F_u(\boldsymbol{x}_u) &= \int_{[0,1]^{\bar{u}}} \left[F(\boldsymbol{x}) - \sum_{v \subset u} F_v(\boldsymbol{x}_v) \right] d\boldsymbol{x}_{\bar{u}}. \end{aligned}$$

Here, $[0,1]^u$ denotes the space of values for components of x_r with $r \in u$, $\boldsymbol{x}_u$ denotes the coordinate projection of $\boldsymbol{x}$ onto $[0,1]^u$. Putting independent priors on F_u, $u \subset \{1,\ldots,s\}$ as follows:

$$F_\emptyset \sim N(0,1),$$

$$F_u(\boldsymbol{x}_u) \sim B_u(\boldsymbol{x}_u) - \sum_{v \subset u} \int_{[0,1]^v} B_u(\boldsymbol{x}_u) d\boldsymbol{x}_v,$$

where $B_u(\boldsymbol{x}_u)$ is a Brownian sheet with variance $\theta^{|u|}$ where $|u|$ is the cardinality of u. It follows that the covariance kernel of F is $K_3(\boldsymbol{x},\boldsymbol{t})$ (Barry, 1986).

3. Optimal designs

In this section we find the design that minimizes $\Psi(\xi;K_j)$ for $j=1,2,3$ respectively, which have N runs and s q-level factors. First we introduce some notation. According to Ma et al. (2003), we define

$$\mathcal{G} = \{(l_1,\ldots,l_s) \mid l_r \in \{1,\ldots,q\},\ r=1,\ldots,s\},$$

We consider the design $\xi = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$, where each run $\boldsymbol{x}_i$ is of the lattice form

$$\boldsymbol{x}_i = \left(\frac{2l_1-1}{2q},\ldots,\frac{2l_s-1}{2q}\right), \text{ where } (l_1,\ldots,l_s) \in \mathcal{G}. \tag{4}$$

Given N, s and q, denote $\Xi(N;q^s)$ as the set of all such ξ. Note that $\mathcal{U}(N;q^s) \subset \mathcal{G}(N;q^s)$.

For a given design $\xi \in \Xi(N;q^s)$, denote by $n_{(l_1,\ldots,l_s)}$ the number of runs at the level-combination $(l_1,\ldots,l_s) \in \mathcal{G}$, and let $\boldsymbol{n}_\xi$ be a q^s-vector with elements $n_{(l_1,\ldots,l_s)}$ arranged lexicographically. For a given $\xi \in \Xi(N;q^s)$, $\boldsymbol{n}_\xi/N$ can be regarded as a probability measure over q^s level-combinations. Therefore, extend $\boldsymbol{n}_\xi$ to be a measure on q^s level-combinations.

3.1 The case with covariance kernel K_1

We first consider the criterion $\Psi(\xi;K_1)$. Note that

$$\int_{[0,1]^s} K_1(\boldsymbol{x},\boldsymbol{x})d\boldsymbol{x} = \frac{1}{6^s}, \quad \int_{[0,1]^s} K_1(\boldsymbol{x},\boldsymbol{x}_i)d\boldsymbol{x} = \prod_{r=1}^s \left[\frac{1}{8} - \frac{1}{2}\left|x_{ir} - \frac{1}{2}\right|^2\right].$$

Then the quantity $\Psi(\xi;K_1)$ is

$$\Psi(\xi;K_1) = \frac{1}{6^s} - 2\hat{\gamma}\sum_{i=1}^N \prod_{r=1}^s \left[\frac{1}{8} - \frac{1}{2}\left|x_{ir} - \frac{1}{2}\right|^2\right] + \hat{\gamma}^2 \sum_{i,j=1}^N \prod_{r=1}^s [\min(x_{ir},x_{jr}) - x_{ir}x_{jr}]. \tag{5}$$

For the lattice design ξ with N-runs of the form (4), we define

$$a_l = \frac{1}{8} - \frac{1}{2}\left|\frac{2l-1-q}{2q}\right|^2, \quad a_{kl} = \frac{2\min(k,l)-1}{2q} - \frac{(2k-1)(2l-1)}{4q^2}, \quad k,l=1,\ldots,q.$$

Then $\Psi(\xi; K_1)$ in (5) can be expressed as follows:

$$\Psi(\xi;K_1) = \frac{1}{6^s} - 2\hat{\gamma} \sum_{(k_1,\ldots,k_s)\in\mathcal{G}} n_{(k_1,\ldots,k_s)} \prod_{r=1}^{s} a_{k_r}$$
$$+\hat{\gamma}^2 \sum_{(k_1,\ldots,k_s)\in\mathcal{G}} \sum_{(l_1,\ldots,l_s)\in\mathcal{G}} n_{(k_1,\ldots,k_s)} n_{(l_1,\ldots,l_s)} \prod_{r=1}^{s} a_{k_r l_r}.$$

Define $\boldsymbol{a}_0 = (a_1, \ldots, a_s)'$, $\boldsymbol{A}_0 = (a_{kl})_{q\times q}$ and $\boldsymbol{a}_s = \otimes^s \boldsymbol{a}_0$, $\boldsymbol{A}_s = \otimes^s \boldsymbol{A}_0$, where $\otimes$ is the Kronecker product. By the definition of q^s-vector $\boldsymbol{n}_\xi$ we can further express $\Psi(\xi; K_1)$ in the following form

$$\Psi(\xi; K_1) = \frac{1}{6^s} - 2\hat{\gamma}\boldsymbol{a}_s'\boldsymbol{n}_\xi' + \hat{\gamma}\boldsymbol{n}_\xi'\boldsymbol{A}_s\boldsymbol{n}_\xi. \tag{6}$$

Moreover, the matrix $\boldsymbol{A}_0$ has the following properties, which is useful in proving Theorem 1.

Lemma 1. *Let $\boldsymbol{a}_0$ and $\boldsymbol{A}_0$ defined above. Then we have*

(1) $\boldsymbol{A}_0^{-1}$ is a $q\times q$ tridiagonal symmetric matrix. Its diagonal elements are $(3q,\ 2q\mathbf{1}_{q-2}',\ 3q])'$, *its all second diagonal elements are $-q$.*

(2) $\boldsymbol{A}_0^{-1}\mathbf{1}_q = (2q,\ q\mathbf{1}_{q-2},\ 2q)'$.

(3) $\mathbf{1}_q'\boldsymbol{A}_0^{-1}\mathbf{1}_q = q(q+2)$.

(4) $\boldsymbol{A}_0^{-1}\boldsymbol{a}_0 = (3/(4q),\ (q^2+1)/(8q)\mathbf{1}_{q-2}',\ 3/(4q))'$.

(5) $\mathbf{1}_q'\boldsymbol{A}_0^{-1}\boldsymbol{a}_0 = (q^3 - 2q^2 + q + 10)/(8q)$.

Theorem 1. *A lattice design ξ minimizes $\Psi(\xi; K_1)$ over $\Xi(N; q^s)$ if and only if $\boldsymbol{n}_\xi$ is of the following form:*

$$\boldsymbol{n}_\xi = \frac{1}{\hat{\gamma}(8q)^s}\Big\{(q^2+1)^s \otimes^s \begin{pmatrix} 6/(q^2+1) \\ \mathbf{1}_{q-2} \\ 6/(q^2+1) \end{pmatrix} + \frac{N\hat{\gamma}(8q)^s - (q^3-2q^2+q+10)^s}{(q+2)^s} \otimes^s \begin{pmatrix} 2 \\ \mathbf{1}_{q-2} \\ 2 \end{pmatrix}\Big\}. \tag{7}$$

In particular, for $q = 2$ the design ξ is a complete design with $\boldsymbol{n}_\xi = (N/2^s)\mathbf{1}_{2^s}$.

Proof. From (6) and the Lagrange multiplier method, let

$$L(\boldsymbol{n}_\xi, \lambda) = \frac{1}{6^s} - 2\hat{\gamma}\boldsymbol{a}_s'\boldsymbol{n}_\xi' + \hat{\gamma}\boldsymbol{n}_\xi'\boldsymbol{A}_s\boldsymbol{n}_\xi + \lambda\left(\boldsymbol{n}_\xi'\mathbf{1}_{q^s} - N\right).$$

The following system of equations

$$\begin{cases} \dfrac{\partial L}{\partial \lambda} = \boldsymbol{n}_\xi'\mathbf{1}_{q^s} - N = 0, \\ \dfrac{\partial L}{\partial \boldsymbol{n}_\xi} = -2\hat{\gamma}\boldsymbol{a}_s + 2\hat{\gamma}^2\boldsymbol{A}_s\boldsymbol{n}_\xi + \lambda\mathbf{1}_{q^s} = \mathbf{0} \end{cases}$$

gives

$$\lambda = \frac{2\hat{\gamma}\mathbf{1}'_{q^s}\boldsymbol{A}_s^{-1}\boldsymbol{a}_s - 2N\hat{\gamma}^2}{\mathbf{1}'_{q^s}\boldsymbol{A}_s^{-1}\mathbf{1}_{q^s}}, \tag{8}$$

and

$$\boldsymbol{n}_\xi = \frac{1}{\hat{\gamma}}\boldsymbol{A}_s^{-1}\boldsymbol{a}_s - \frac{1}{2\hat{\gamma}^2}\lambda\boldsymbol{A}_s^{-1}\mathbf{1}_{q^s}. \tag{9}$$

From Lemma 1 we find that

$$\begin{aligned}\boldsymbol{A}_s^{-1}\mathbf{1}_{q^s} &= q^s \otimes^s (2, \mathbf{1}'_{q-2}, 2)', \\ \boldsymbol{A}_s^{-1}\boldsymbol{a}_s &= \left(\frac{q^2+1}{8q}\right)^s \otimes^s (6/(q^2+1), \mathbf{1}'_{q-2}, 6/(q^2+1))',\end{aligned}$$

and

$$\mathbf{1}'_{q^s}\boldsymbol{A}_s^{-1}\mathbf{1}_{q^s} = [q(q+2)]^s, \quad \mathbf{1}'_{q^s}\boldsymbol{A}_s^{-1}\boldsymbol{a}_s = \left(\frac{q^3-2q^2q+10}{8q}\right)^s.$$

The result (12) follows from making use of these facts in (8) and (9). When $q = 2$, (12) becomes $\boldsymbol{n}_\xi = (N/2^s)\mathbf{1}_{2^s}$. ∎

3.2 The case with covariance kernel K_2

We now consider the criterion $\Psi(\xi; K_2)$. Note that

$$\int_{[0,1]^s} K_2(\boldsymbol{x}, \boldsymbol{x})d\boldsymbol{x} = 1, \quad \int_{[0,1]^s} K_2(\boldsymbol{x}, \boldsymbol{x}_i)d\boldsymbol{x} = \prod_{r=1}^{s}\left[\frac{1}{4} - \left|x_{ir} - \frac{1}{2}\right|^2\right].$$

Then the quantity $\Psi(\xi; K_2)$ is

$$\Psi(\xi; K_2) = 1 - 2\hat{\gamma}\sum_{i=1}^{N}\prod_{r=1}^{s}\left[\frac{1}{4} - \left|x_{ir} - \frac{1}{2}\right|^2\right] + \hat{\gamma}^2\sum_{i,j=1}^{N}\prod_{r=1}^{s}[1 - |x_{ir} - x_{jr}|]. \tag{10}$$

For the lattice design ξ with N-runs of the form (4), define

$$b_l = \frac{1}{4} - \left|\frac{2l-1-q}{2q}\right|^2, \quad b_{kl} = 1 - \left|\frac{k-l}{q}\right|, \quad k, l = 1, \ldots, q,$$

and $\boldsymbol{b}_0 = (b_1, \ldots, b_s)'$, $\boldsymbol{B}_0 = (b_{kl})_{q\times q}$, $\boldsymbol{b}_s = \otimes^s\boldsymbol{b}_0$, $\boldsymbol{B}_s = \otimes^s\boldsymbol{B}_0$. we can further express $\Psi(\xi; K_2)$ in the following form

$$\Psi(\xi; K_2) = 1 - 2\hat{\gamma}\boldsymbol{b}'_s\boldsymbol{n}'_\xi + \hat{\gamma}\boldsymbol{n}'_\xi\boldsymbol{B}_s\boldsymbol{n}_\xi. \tag{11}$$

The matrix $\boldsymbol{B}_0$ has the following properties.

Lemma 2. *Let $\boldsymbol{b}_0$ and $\boldsymbol{B}_0$ defined above. Then we have*

(1) $\boldsymbol{B}_0^{-1}$ is a $q\times q$ symmetric matrix. Its diagonal elements are $(q(q+2)/[2(q+1)],\ q\mathbf{1}'_{q-2},\ q(q+2)/[2(q+1)])'$, its all second diagonal elements are $-q/2$, its $(1,q)$- and $(q,1)$-elements are $q/[2(q+1)]$, and all others are 0.

(2) $\boldsymbol{B}_0^{-1}\mathbf{1}_q = (q/(q+1),\ \mathbf{0}_{q-2},\ q/(q+1))'$.

(3) $\mathbf{1}_q'\boldsymbol{B}_0^{-1}\mathbf{1}_q = 2q/(q+1)$.

(4) $\boldsymbol{B}_0^{-1}\boldsymbol{b}_0 = ((-2q^2+2q+3)/[4q(q+1)],\ q^{-1}\mathbf{1}_{q-2}',\ (-2q^2+2q+3)/[4q(q+1)])'$.

(5) $\mathbf{1}_q'\boldsymbol{B}_0^{-1}\boldsymbol{b}_0 = (2q-1)/[2q(q+1)]$.

Making use of this lemma and the way in the proof of Theorem 1 we can prove the following theorem:

Theorem 2. *A lattice design ξ minimizes $\Psi(\xi; K_2)$ over $\Xi(N; q^s)$ if and only if $\boldsymbol{n}_\xi$ is of the following form:*

$$\boldsymbol{n}_\xi = \frac{1}{\hat{\gamma}[4q(q+1)]^s}\Big\{\otimes^s \begin{pmatrix} -2q^2+4q+3 \\ 4(q+1)\mathbf{1}_{q-2} \\ -2q^2+4q+3 \end{pmatrix} + [N\hat{\gamma}(2q(q+1))^s - (2q-1)^s]\otimes^s \begin{pmatrix} 1 \\ \mathbf{0}_{q-2} \\ 1 \end{pmatrix}\Big\}. \quad (12)$$

In particular, for $q = 2$ the design ξ is a complete design with $\boldsymbol{n}_\xi = (N/2^s)\mathbf{1}_{2^s}$.

3.3 The case with covariance kernel K_3

We note that $K_3(\boldsymbol{x}, \boldsymbol{t})$ can be expressed by

$$K_3(\boldsymbol{x}, \boldsymbol{t}) = \prod_{r=1}^{s}\left[1 + \frac{\theta}{2}\left(\frac{1}{6} + \left|x_r - \frac{1}{2}\right|^2 + \left|t_r - \frac{1}{2}\right|^2 - |x_r - t_r|\right)\right],$$

and

$$\int_{[0,1]^s} K_3(\boldsymbol{x}, \boldsymbol{x})d\boldsymbol{x} = \left(1 + \frac{\theta}{6}\right)^s, \quad \int_{[0,1]^s} K_3(\boldsymbol{x}, \boldsymbol{x}_i)d\boldsymbol{x} = 1.$$

Then the quantity $\Psi(\xi; K_3)$ is

$$\Psi(\xi; K_3) = \left(1 + \frac{\theta}{6}\right)^s - 2N\hat{\gamma} + \hat{\gamma}^2 \sum_{i,j=1}^{N}\prod_{r=1}^{s}\left[1 + \frac{\theta}{2}\left(\frac{1}{6} + \left|x_{ir} - \frac{1}{2}\right|^2 + \left|x_{jr} - \frac{1}{2}\right|^2 - |x_{ir} - x_{jr}|\right)\right]. \quad (13)$$

For the lattice design ξ with N-runs of the form (4), we define

$$c_{kl} = 1 + \frac{\theta}{2}\left(\frac{1}{6} + \left|\frac{2k-1}{2q} - \frac{1}{2}\right|^2 + \left|\frac{2l-1}{2q} - \frac{1}{2}\right|^2 - \left|\frac{k-l}{q}\right|\right), \quad k, l = 1, \ldots, q,$$

and $\boldsymbol{C}_0 = (c_{kl})_{q\times q}$, $\boldsymbol{C}_s = \otimes^s \boldsymbol{C}_0$. It follows that

$$\Psi(\xi; K_3) = \left(1 + \frac{\theta}{6}\right)^s - 2N\hat{\gamma} + \hat{\gamma}\boldsymbol{n}_\xi'\boldsymbol{C}_s\boldsymbol{n}_\xi. \quad (14)$$

Theorem 3. *A lattice design ξ minimizes $\Psi(\xi; K_3)$ over $\Xi(N; q^s)$ if and only if $\boldsymbol{n}_\xi$ is of the following form:*

$$\boldsymbol{n}_\xi = \frac{N}{\mathbf{1}_{q^s}'\boldsymbol{C}_s^{-1}\mathbf{1}_{q^s}}\boldsymbol{C}_s^{-1}\mathbf{1}_{q^s}.$$

In particular, for $q = 2$ the design ξ is a complete design with $\boldsymbol{n}_\xi = (N/2^s)\mathbf{1}_{2^s}$.

Proof. From (14) and the Lagrange multiplier method, let

$$L(\boldsymbol{n}_\xi, \lambda) = \left(1 + \frac{\theta}{6}\right)^s - 2N\hat{\gamma} + \hat{\gamma}\boldsymbol{n}_\xi' \boldsymbol{C}_s \boldsymbol{n}_\xi + \lambda\left(\boldsymbol{n}_\xi'\mathbf{1}_{q^s} - N\right).$$

The following system of equations

$$\begin{cases} \dfrac{\partial L(\boldsymbol{n}_\xi, \lambda)}{\partial \lambda} = \boldsymbol{n}_\xi'\mathbf{1}_{q^s} - N = 0, \\ \dfrac{\partial L(\boldsymbol{n}_\xi, \lambda)}{\partial \boldsymbol{n}_\xi} = 2\hat{\gamma}^2 \boldsymbol{C}_s \boldsymbol{n}_\xi + \lambda \mathbf{1}_{q^s} = \mathbf{0} \end{cases}$$

gives $\lambda = -2N\hat{\gamma}^2/\mathbf{1}_{q^s}'\boldsymbol{C}_s^{-1}\mathbf{1}_{q^s}$, and

$$\boldsymbol{n}_\xi = -\frac{1}{2\hat{\gamma}^2}\lambda \boldsymbol{C}_s^{-1}\mathbf{1}_{q^s} = \frac{N}{\mathbf{1}_{q^s}'\boldsymbol{C}_s^{-1}\mathbf{1}_{q^s}}\boldsymbol{C}_s^{-1}\mathbf{1}_{q^s}.$$

For the case with $q = 2$, we find that

$$\boldsymbol{C}_0 = \frac{1}{48}\begin{pmatrix} 48 + 7\theta & 48 - 5\theta \\ 48 - 5\theta & 48 + 7\theta \end{pmatrix}, \quad \boldsymbol{C}_0^{-1} = \frac{2}{\theta(48 + \theta)}\begin{pmatrix} 48 + 7\theta & -48 + 5\theta \\ -48 + 5\theta & 48 + 7\theta \end{pmatrix},$$

and

$$\boldsymbol{C}_s^{-1}\mathbf{1}_{2^s} = \otimes^s\left(\boldsymbol{C}_0^{-1}\mathbf{1}_2\right) = \left(\frac{24}{48 + \theta}\right)^s \mathbf{1}_{2^s}.$$

Straightforward calculation yields that $\boldsymbol{n}_\xi = (N/2^s)\mathbf{1}_{2^s}$. ∎

4. Alternative Formulations and Lower bounds for Ψ

In this section we give some alternative formulations and lower bounds for $\Psi(\xi; K_j)$, $j = 1, 2, 3$, respectively, by using the approach used in Fang et al. (2003).

4.1 Based on the column balance

Let $\boldsymbol{U}_{N\times s}$ be the matrix of a U-type design $U(N, q^s)$, whose elements are integers $1, \ldots, q$. and let $\boldsymbol{X}_{N\times s}$ be the matrix induced form $\boldsymbol{U}_{N\times s}$ by mapping $u_{ir} \to x_{ir} = (2u_{ir} - 1)/(2q)$. Denote by $\boldsymbol{u}^r$ the r-th column of $\boldsymbol{U}$, $r = 1, \ldots, s$. For each m columns of $\boldsymbol{U}$, say, $\boldsymbol{u}^{c_1}, \ldots, \boldsymbol{u}^{c_m}$, define

$$B_{c_1,\ldots,c_m} = \sum_{1\le l_1,\ldots,l_m\le q}\left(n_{l_1,\ldots,l_m}^{(c_1,\ldots,c_m)} - \frac{N}{q^m}\right)^2, \quad B(m) = \sum_{1\le c_1,\ldots,c_m\le s} B_{c_1,\ldots,c_m}\Big/\binom{s}{m}, \qquad (15)$$

where $n_{l_1,\ldots,l_m}^{(c_1,\ldots,c_m)}$ is the number of rows in which the column group $\{\boldsymbol{u}^{c_1}, \ldots, \boldsymbol{u}^{c_m}\}$ takes the level combination $\{l_1, \ldots, l_m\}$, and the summation is taken over all possible level combinations.

$B(m)$ measures the closeness to orthogonality of strength m of the design $\boldsymbol{U}$. Fang et al. (2003) showed the following result:

Lemma 3. *Let $\boldsymbol{U}_{N\times s}$ be the matrix of a U-type design $U(N,q^s)$. Then*

$$B(m)=\sum_{i,j=1}^{N}\sum_{1\le c_1,\ldots,c_m\le s}\delta_{ij}^{(c_1,\ldots,c_m)}\Big/\binom{s}{m}-\frac{N^2}{q^m}, \tag{16}$$

where

$$\delta_{ij}^{(c_1,\ldots,c_m)}=\begin{cases}1, & \text{when } x_{ir}=x_{jr} \text{ for } r=c_1,\ldots,c_m,\\ 0, & \text{otherwise.}\end{cases} \tag{17}$$

For two-level U-type designs in $\mathcal{U}(N;2^s)$, $\Psi(\xi,K_j)$, $j=1,2,3$, can be expressed as functions of $(B(1),\ldots,B(s))$.

Theorem 4. *For a two-level design $\xi\in\mathcal{U}(N,2^s)$ with the design matrix $\boldsymbol{U}$ and induced matrix $\boldsymbol{X}$, we have*

$$\Psi(\xi;K_1)=\frac{1}{6^s}-2N\hat{\gamma}\left(\frac{3}{32}\right)^s+N^2\hat{\gamma}^2\left(\frac{1}{8}\right)^s+\frac{\hat{\gamma}^2}{16^s}\sum_{m=1}^{s}2^m\binom{s}{m}B(m), \tag{18}$$

$$\Psi(\xi;K_2)=1-2N\hat{\gamma}\left(\frac{3}{16}\right)^s+N^2\hat{\gamma}^2\left(\frac{3}{4}\right)^s+\frac{\hat{\gamma}^2}{2^s}\sum_{m=1}^{s}\binom{s}{m}B(m), \tag{19}$$

and

$$\Psi(\xi;K_3)=\left(1+\frac{\theta}{6}\right)^s-2N\hat{\gamma}+N^2\hat{\gamma}^2\left(1+\frac{\theta}{48}\right)^s+\hat{\gamma}^2\left(1-\frac{5\theta}{48}\right)^s\sum_{m=1}^{s}\left(\frac{12\theta}{48-5\theta}\right)^m\binom{s}{m}B(m). \tag{20}$$

Proof. For two-level case, i.e., $q=2$, the values of x_{ir}, the elements of $\boldsymbol{X}$, are $\frac{1}{4}$ and $\frac{3}{4}$. Therefore, we have

$$\left|x_{ir}-\frac{1}{2}\right|=\frac{1}{4},\quad \min(x_{ir},x_{jr})-x_{ir}x_{jr}=\frac{1}{16}\left(1+2\delta_{ij}^{(r)}\right), \tag{21}$$

where $\delta_{ij}^{(r)}$ is as defined in (17). Substituting (21) into (10) and making use of the definition of $\delta_{ij}^{(c_1,\ldots,c_m)}$ yields

$$\begin{aligned}\Psi(\xi;K_1)&=\frac{1}{6^s}-2N\hat{\gamma}\left(\frac{3}{32}\right)^s+\frac{\hat{\gamma}^2}{16^s}\sum_{i,j=1}^{N}\prod_{r=1}^{s}\left(1+2\delta_{ij}^{(r)}\right)\\&=\frac{1}{6^s}-2N\hat{\gamma}\left(\frac{3}{32}\right)^s+\frac{\hat{\gamma}^2}{16^s}\left[N^2+\sum_{i,j=1}^{N}\sum_{m=1}^{s}2^m\sum_{c_1<\cdots<c_m}\delta_{ij}^{(c_1,\ldots,c_m)}\right].\end{aligned} \tag{22}$$

The result (18) follows from (16) with $q=2$.

The results (19) and (20) can be proved in the same way. The theorem is proved. ∎

The expressions (18), (19) and (20) allow us to obtain lower bounds for the quantities $\Psi(\xi; K_j)$, $j = 1, 2, 3$, respectively, over two-level U-type designs. Note that $B(m)$ in defined in (15) has a lower bound as follows (Fang et al. 2003):

$$B(m) \geq R_{N,m,2}\left(1 - \frac{R_{N,m,2}}{2^m}\right),$$

where $R_{N,m,2}$ is the residual of $N \pmod{2^m}$. We then have the following results:

Theorem 5. *For a two-level U-type design $U(N, 2^s)$, we have*

(1) $\Psi(\xi; K_1) \geq L_1^{\text{col}}$ where

$$L_1^{\text{col}} = \frac{1}{6^s} - 2N\hat{\gamma}\left(\frac{3}{32}\right)^s + \frac{N^2\hat{\gamma}^2}{8^s} + \frac{\hat{\gamma}^2}{16^s}\sum_{m=1}^{s} 2^m \binom{s}{m} R_{N,m,2}\left(1 - \frac{R_{N,m,2}}{2^m}\right). \tag{23}$$

(2) $\Psi(\xi; K_2) \geq L_2^{\text{col}}$ where

$$L_2^{\text{col}} = 1 - 2N\hat{\gamma}\left(\frac{3}{16}\right)^s + N^2\hat{\gamma}^2\left(\frac{3}{4}\right)^s + \frac{\hat{\gamma}^2}{2^s}\sum_{m=1}^{s}\binom{s}{m} R_{N,m,2}\left(1 - \frac{R_{N,m,2}}{2^m}\right). \tag{24}$$

(3) $\Psi(\xi; K_3) \geq L_3^{\text{col}}$ where

$$\begin{aligned} L_3^{\text{col}} = & \left(1 + \frac{\theta}{6}\right)^s - 2N\hat{\gamma} + N^2\hat{\gamma}^2\left(1 + \frac{\theta}{48}\right)^s \\ & + \hat{\gamma}^2\left(1 - \frac{5\theta}{48}\right)^s \sum_{m=1}^{s}\left(\frac{12\theta}{48 - 5\theta}\right)^m \binom{s}{m} R_{N,m,2}\left(1 - \frac{R_{N,m,2}}{2^m}\right). \end{aligned} \tag{25}$$

4.2 Based on the row distance

Let ξ be a design $U(N, q^s)$, and let $\boldsymbol{X} = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_N')'$ be the corresponding design matrix, where $\boldsymbol{x}_i$ is its i-th row, $i = 1, \ldots, n$. According to Fang et al. (2003), let d_{ij} be the number of columns for which $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ take different values. Then $\lambda_{ij} = s - d_{ij}$ is the number of columns for which $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ take the same value. The set $\{\lambda_{ij},\ 1 \leq i < j \leq n\}$ characterizes the relation between the lows of the design $\boldsymbol{X}$. It is known that

$$\sum_{j \neq i} \lambda_{ij} = s\left(\frac{N}{q} - 1\right), \quad i = 1, \ldots, N, \tag{26}$$

and

$$\sum_{c_1 < \cdots < c_m} \delta_{ij}^{(c_1, \ldots, c_m)} = \binom{\lambda_{ij}}{m}, \tag{27}$$

where $\delta_{ij}^{(c_1, \ldots, c_m)}$ is defined as in (17).

We can express $\Psi(\xi; K_j)$, $j = 1, 2, 3$, in terms of the λ_{ij} as follows:

Theorem 6. *For a two-level U-type design $\xi \in \mathcal{U}(N, 2^k)$ with the design matrix $\boldsymbol{X}$, we have*

$$\Psi(\xi; K_1) = \frac{1}{6^s} - 2N\hat{\gamma}\left(\frac{3}{32}\right)^s + \frac{\hat{\gamma}^2}{16^s}\sum_{i,j=1}^{N} 3^{\lambda_{ij}}, \tag{28}$$

$$\Psi(\xi; K_2) = 1 - 2N\hat{\gamma}\left(\frac{3}{16}\right)^s + \frac{\hat{\gamma}^2}{2^s}\sum_{i,j=1}^{N} 2^{\lambda_{ij}}, \tag{29}$$

$$\Psi(\xi; K_3) = \left(1 + \frac{\theta}{6}\right)^s - 2N\hat{\gamma} + \hat{\gamma}^2\left(1 - \frac{5\theta}{48}\right)^s \sum_{i,j=1}^{N}\left(\frac{48 + 7\theta}{48 - 5\theta}\right)^{\lambda_{ij}}. \tag{30}$$

Proof. Substituting (27) into (22) yields the expression (28). In the similar way, we can prove (29) and (30). ∎

The formulations of the $\Psi(\xi; K_j)$, $j = 1, 2, 3$, in this theorem allow us to derive lower bounds different from those in Theorem 4 by approach used in Fang et al. (2003).

Theorem 7. *Let ξ be a two-level U-type design in $\mathcal{U}(N, 2^s)$. Define $\lambda = s(N-2)/[2(N-1)]$. We then have*

(1) $\Psi(\xi; K_1) \geq L_1^{\text{row}}$ where

$$L_1^{\text{row}} = \frac{1}{6^s} - 2N\hat{\gamma}\left(\frac{3}{32}\right)^s + N\hat{\gamma}^2\left(\frac{3}{16}\right)^s + \frac{\hat{\gamma}^2}{16^s}N(N-1)3^{\lambda}. \tag{31}$$

(2) $\Psi(\xi; K_2) \geq L_2^{\text{row}}$ where

$$L_2^{\text{row}} = 1 - 2N\hat{\gamma}\left(\frac{3}{16}\right)^s + N\hat{\gamma}^2 + \frac{\hat{\gamma}^2}{2^s}N(N-1)2^{\lambda}. \tag{32}$$

(3) $\Psi(\xi; K_3) \geq L_3^{\text{row}}$ where

$$L_3^{\text{row}} = \left(1 + \frac{\theta}{6}\right)^s - 2N\hat{\gamma} + N\hat{\gamma}^2\left(1 + \frac{7\theta}{48}\right)^s + \hat{\gamma}^2\left(1 - \frac{5\theta}{48}\right)^s N(N-1)\left(\frac{48 + 7\theta}{48 - 5\theta}\right)^{\lambda}. \tag{33}$$

Proof. Let Y be a random variable which is uniformly distributed on the set $\{\lambda_{ij},\ 1 \leq i < j \leq N\}$, and then

$$E(Y) = s(N-2)/[2(N-1)] = \lambda,$$

according to (26).

To Prove (31), we define $g(y) = 3^y$, which is a convex function of y. By Jenssen's inequality we then have

$$E[g(Y)] = \frac{1}{N(N-1)}\sum_{1 \leq i \neq j \leq N} 2^{\lambda_{ij}} \geq g(E[Y]) = 2^{\lambda},$$

and the lower bound (31) follows from (28) by noting that $\lambda_{ii} = s$. In the similar way, we can prove lower bounds (32) and (33). ∎

5. Summary

In this paper, we have considered the design problem for recovering a response surface by using nonparametric Bayesian approach. The criterion for choosing design is developed based on the average estimation variance by asymptotic used in Mitchell et al. (1994). Three priors for the response are considered.

For each of the three priors, we have found the optimal design that minimizes the criterion over the lattice designs with s q-level factors and N runs. The approach we used is similar to that in Ma et al. (2003). It is shown that for the case with $q = 2$ the complete design is optimal for each of the three Bayes models.

We also obtained alternative expressions and lower bounds for the criterion corresponding to each of the three Bayes models for the two-level U-type design by using the column balance and row distance proposed in Fang et al. (2003). These expressions and lower bounds can be useful in searching the two-level U-type designs for the Bayes models. This will be our further work to be done.

References

[1] Barry, D. (1986). Nonparametric Bayesian regression, *Ann. Statist.* **14**, 934-953.

[2] Hickernell, F. J. (1998). A generalized discrepancy and quadrature error bound. *Math. Comput.* **67**, 299-322.

[3] Ma Chang-Xing, Fang Kai-Tai, Lin Dennis K. J. (2003). A note on uniformity and orthogonality . *Jounal of stastical planning and inference.* **113**, 323-334.

[4] Fang Kai-Tai, Lu Xuan and Winker P. (2003). Lower bounds for centered and wrap-around L_2-discrepancies and construction of uniform designs by threshold accepting. *Journal of Complexity* to appear.

[5] Mitchell, T., J. Sacks and D. Ylvisaker (1994). Asymptotic Bayes criteria for nonparametric response surface design, *Ann. Statist.* **22**, 634-651.

[6] Yue Rong-Xian (2001). A comparison of random and quasirandom points for nonparametric response surface design. *Statistics and Probability Letters*, **53**, 129-142.

[7] Steinberg, D.M. (1985). Model robust response surface designs: Scaling two-level factorials, *Biometrika* **72**, 513-526.