A Study on Analytical Properties of Bayesian Experimental Design Model based on an Orthonormal System

Yoshifumi Ukita[1] Shunsuke Horii[2] Toshiyasu Matsushima[2]

[1]Yokohama College of Commerce

[2]Waseda University

Bayes on the Beach 2017
November 13th -15th 2017

Acknowledgement: JSPS KAKENHI Grant Number JP17K00316

Outline

1. Introduction
   - Motivation and Objective
   - Orthonormal Systems
2. Experimental Design Models Based on Orthonormal Systems
   - Previous model
   - Model based on orthonormal systems
   - Relation between the two models
3. Bayesian Experimental Designs
   - Prior and Posterior Probability
   - Posterior variance for orthogonal designs
4. Conclusion

Motivation and Objective

In signal processing, the model is generally based on orthonormal systems. In such models

- All parameters are independent.
- The fast Fourier Transform (FFT) can calculate the parameters.
- Parameters are complex numbers.

In some previous experimental design models

- Parameters are not necessarily independent.
- Parameters are real numbers.

Motivation

Can we use the idea of orthonormal systems to calculate the posterior variance in Bayesian experimental designs?

Objective

We focus on a subclass of designs, which is limited by orthonormal systems. Then, it's to show we can get the posterior variance directly.

Orthonormal Systems: Example 1

$F_1, F_2, \ldots, F_n$: Factors
$x_i \in \{0,1\}$: the level of F_i
$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n$: level combination
$\{0,1\}^n$: the set of all sequences of $0, 1$ that is n long

Basis functions over the Boolean domain

For each $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \{0,1\}^n$, define the basis function

$$\mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) = (-1)^{\boldsymbol{a} \cdot \boldsymbol{x}},$$

where $\boldsymbol{a} \cdot \boldsymbol{x} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$ is to be performed modulo 2.

Orthonormal Systems

The basis functions form orthonormal systems, that is,

$$\frac{1}{2^n} \sum_{\boldsymbol{x} \in \{0,1\}^n} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) \mathcal{X}_{\boldsymbol{b}}(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{a} = \boldsymbol{b}, \\ 0, & \boldsymbol{a} \neq \boldsymbol{b}. \end{cases}$$

Orthonormal Systems: Example 2

$F_1, F_2, \ldots, F_n$: Factors
$x_i \in \{0,1,2\}$: the level of F_i
$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \{0,1,2\}^n$: level combination
$\{0,1,2\}^n$: the set of all sequences of $0,1,2$ that is n long

Basis functions over $\{0,1,2\}^n$ domain ($GF(3)^n$ domain)

For each $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \{0,1,2\}^n$, define the basis function

$$\mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) = e^{i2\pi \boldsymbol{a}\cdot\boldsymbol{x}/3},$$

where $\boldsymbol{a}\cdot\boldsymbol{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ is to be performed over $GF(3)$.

Orthonormal Systems

The basis functions form orthonormal systems, that is,

$$\frac{1}{3^n} \sum_{\boldsymbol{x}\in\{0,1,2\}^n} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x})\mathcal{X}^*_{\boldsymbol{b}}(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{a} = \boldsymbol{b}, \\ 0, & \boldsymbol{a} \neq \boldsymbol{b}, \end{cases}$$

where $\mathcal{X}^*_{\boldsymbol{b}}(\boldsymbol{x})$ is the complex conjugate of $\mathcal{X}_{\boldsymbol{b}}(\boldsymbol{x})$.

Example: Model of Two-factor Experiments

F_1, F_2 : the factors
$x_i \in \{0, 1, 2\}$: the level of F_i
$\boldsymbol{x} = (x_1, x_2) \in \{0, 1, 2\}^2$: level combination
$t(\boldsymbol{x})$:the response of the experiment with level combination $\boldsymbol{x}$

Previous model

$$t(\boldsymbol{x}) = \mu + \alpha_1(x_1) + \alpha_2(x_2) + \beta_{1,2}(x_1, x_2) + \epsilon,$$

where
μ: the general mean
$\alpha_1(x_1)$: the effect of the x_1 level of F_1
$\alpha_2(x_2)$: the effect of the x_2 level of F_2
$\beta_{1,2}(x_1, x_2)$: the interaction of the x_1 level of F_1 and the x_2 level of F_2
ϵ: a zero-mean Gaussian random variable with variance σ^2

Parameters on Previous Model

Then, all parameters are given as follows:
$\mu, \alpha_1(0), \alpha_1(1), \alpha_1(2), \alpha_2(0), \alpha_2(1), \alpha_2(2), \beta_{1,2}(0,0), \beta_{1,2}(0,1), \beta_{1,2}(0,2),$
$\beta_{1,2}(1,0), \beta_{1,2}(1,1), \beta_{1,2}(1,2), \beta_{1,2}(2,0), \beta_{1,2}(2,1), \beta_{1,2}(2,2).$

independent parameters vector $\boldsymbol{u} \in \mathbb{R}^9$

Under the constraints on parameters, let $\boldsymbol{u}$ denote the independent parameters vector.

$$\boldsymbol{u} = \begin{bmatrix} \mu \\ \alpha_1(0) \\ \alpha_1(1) \\ \alpha_2(0) \\ \alpha_2(1) \\ \beta_{1,2}(0,0) \\ \beta_{1,2}(0,1) \\ \beta_{1,2}(1,0) \\ \beta_{1,2}(1,1) \end{bmatrix},$$

Model based on orthonormal systems

Model based on orthonormal systems (Ukita et al. 2010)

$$t(\boldsymbol{x}) = \sum_{\boldsymbol{a}\in\{0,1,2\}^2} f_{\boldsymbol{a}} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) + \epsilon,$$

where

$f_{\boldsymbol{a}} : f_{\boldsymbol{a}} \in \mathbb{C}^9$

$\mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) = e^{i2\pi \boldsymbol{a}\cdot\boldsymbol{x}/3}$, and

ϵ: a zero-mean Gaussian random variable with variance σ^2

Then, the basis fuctions $\{\mathcal{X}_{\boldsymbol{a}} | \boldsymbol{a}\in\{0,1,2\}^2\}$ form orthonormal systems.

Parameters (Model based on orthonormal systems)

independent parameters vector $\boldsymbol{w} \in \mathbb{C}^9$ (Fourier coefficients vector)

$$\boldsymbol{w} = \begin{bmatrix} f_{00} \\ f_{10} \\ f_{20} \\ f_{01} \\ f_{02} \\ f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \end{bmatrix}.$$

no constraints on parameters $\rightarrow$ all parameters are independent.

Relation between $\boldsymbol{u}$ and $\boldsymbol{w}$

There is a 9×9 matrix $\boldsymbol{M}$ that satisfies

$$\boldsymbol{u} = \boldsymbol{M}\boldsymbol{w},$$

and

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \omega_3 & \omega_3^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_3 & \omega_3^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & \omega_3 & \omega_3^2 & \omega_3 & \omega_3^2 \\ 0 & 0 & 0 & 0 & 0 & \omega_3 & \omega_3 & \omega_3^2 & \omega_3^2 \\ 0 & 0 & 0 & 0 & 0 & \omega_3^2 & 1 & 1 & \omega_3 \end{bmatrix}.$$

where $\omega_3 = e^{2\pi i/3}$. $\boldsymbol{M}^{-1}$ also exists and $\boldsymbol{M}^{-1}\boldsymbol{u} = \boldsymbol{w}$.

Likelihood Function

- a data set of inputs $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ with corresponding target values $t(\boldsymbol{x}_1), \ldots, t(\boldsymbol{x}_N)$
- the variables $\{t(\boldsymbol{x}_1), \ldots, t(\boldsymbol{x}_N)\}$ be a column vector denoted by **t**
- K: the number of pamameters

Likelihood Function

$$p(\mathbf{t}|\boldsymbol{X}, \boldsymbol{u}, \sigma^2) = \mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\boldsymbol{M}^{-1}\boldsymbol{u}, \sigma^2\boldsymbol{I}).$$

where

$$\boldsymbol{\Phi} = \begin{bmatrix} \mathcal{X}_{a_1}(\boldsymbol{x}_1) & \mathcal{X}_{a_2}(\boldsymbol{x}_1) & \ldots & \mathcal{X}_{a_K}(\boldsymbol{x}_1) \\ \mathcal{X}_{a_1}(\boldsymbol{x}_2) & \mathcal{X}_{a_2}(\boldsymbol{x}_2) & \ldots & \mathcal{X}_{a_K}(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}_{a_1}(\boldsymbol{x}_N) & \mathcal{X}_{a_2}(\boldsymbol{x}_N) & \ldots & \mathcal{X}_{a_K}(\boldsymbol{x}_N) \end{bmatrix}.$$

Prior and Posterior Probability

Prior Probability

The corresponding conjugate prior is given by a Gaussian distribution

$$p(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{u}|\boldsymbol{m}_0, \boldsymbol{S}_0).$$

Posterior Probability

Then the posterior probability is given by

$$p(\boldsymbol{u}|\boldsymbol{X}, \mathbf{t}, \sigma^2) = \mathcal{N}(\boldsymbol{u}|\boldsymbol{m}_N, \boldsymbol{S}_N),$$

where

$$\boldsymbol{m}_N = \boldsymbol{S}_N \left(\frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^* \boldsymbol{\Phi}^* \mathbf{t} + \boldsymbol{S}_0^{-1} \boldsymbol{m}_0 \right),$$

$$\boldsymbol{S}_N^{-1} = \frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^* \boldsymbol{\Phi}^* \boldsymbol{\Phi} \boldsymbol{M}^{-1} + \boldsymbol{S}_0^{-1},$$

and $*$ denotes the conjugate transpose (Hermitian transpose).

Except adding $\boldsymbol{M}$, the proof is the same as that of Bishop(2006).

Bayesian experimental designs

There are many criteria for the optimal design (Chaloner & Verdinelli 1995). In this linear model,

Bayesian alphabetic optimality

- A-optimality: Minimize trace $\left[\left(\frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^*\boldsymbol{\Phi}^*\boldsymbol{\Phi}\boldsymbol{M}^{-1} + \boldsymbol{S}_0^{-1}\right)^{-1}\right]$
- D-optimality: Maximize det $\left[\frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^*\boldsymbol{\Phi}^*\boldsymbol{\Phi}\boldsymbol{M}^{-1} + \boldsymbol{S}_0^{-1}\right]$
- et cetera

In this work,

- we focus on a subclass of designs, which is limited by orthonormal systems.

What is the designs which satisfy $\frac{1}{N}\boldsymbol{\Phi}^*\boldsymbol{\Phi} = \boldsymbol{I}$?
$\rightarrow$ Orthogonal designs (Orthogonal arrays), Hedayat et al. (1999)

Posterior variance for orthogonal designs

Posterior variance for orthogonal designs

$$\boldsymbol{S}_N = \left[\left(\frac{N}{\sigma^2} (\boldsymbol{M}^{-1})^* \boldsymbol{M}^{-1} + \boldsymbol{S}_0^{-1} \right)^{-1} \right].$$

Advantages

- Easy to calculate the posterior variance
- the well-balanced design
- Not necessary to search for designs

Disadvantages

- Not depend on the prior distribution
  $\rightarrow$ If the influence of the prior distribution is strong, we should search for the optimal design.
- The number of experiments (inputs) N: restricted to q^k(q: the number of levels of Factor, k: integer)

Conclusion

In this work,

- Focused on a subclass of designs, which satisfies that $\frac{1}{N}\boldsymbol{\Phi}^*\boldsymbol{\Phi} = \boldsymbol{I}$
  $\rightarrow$ orthogonal designs
- Easy to calculate the posterior variance
- Not depend on the prior distribution
  $\rightarrow$ If the influence of the prior distribution is strong, we should search for the optimal design.

Further works

- What is the condition satisfies the orthogonal design is optimal?
- Can we apply this kind of projection to other models?
- Can we use the orthogonal model to search for the optimal design?

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP17K00316.

References

- C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- K. Chaloner and I. Verdinelli, "Bayesian Experimental Design: A Review," *Statistical Science*, Vol.10, No.3, pp.273-304, 1995.
- A.S. Hedayat, N.J.A. Sloane and J. Stufken, *Orthogonal Arrays: Theory and Applications*, Springer, 1999.
- E.M. Stein, R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, 2003.
- H. Toutenburg and Shalabh, *Statistical Analysis of Designed Experiments* (Third Edition), Springer, 2009.
- Y. Ukita, T. Saito, T. Matsushima and S. Hirasawa, "A Note on a Sampling Theorem for Functions over $GF(q)^n$ Domain," *IEICE Trans. Fundamentals*, Vol.E93-A, no.6, pp.1024-1031, June 2010.

Appendix: Orthogonal design 1

In signal processing, the range of frequencies is expressed by the maximum frequency.

In experimental designs, however, it is often necessary to express the range of frequencies in greater detail.

- The range of frequencies can be expressed in greater detail by using a bounded set $A(\subseteq \{0,1\}^n)$ instead of a maximum frequency.

Let A represent all factors that might influence the result of the experiment.

Example

$n = 3, A = \{000, 100, 010, 001, 110\}$.

- 100: factor 1
- 010: factor 2
- 001: factor 3
- 110: interaction of factor 1 and factor 2

The other interactions don't influence the result of the experiment.

Orthogonal design 2

- For $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$, $\boldsymbol{a}' = (a'_1, a'_2, \ldots, a'_n) \in \{0,1\}^n$, define $\boldsymbol{a} + \boldsymbol{a}' = (a_1 + a'_1, a_2 + a'_2, \ldots, a_n + a'_n)$, where $+$ is over $GF(2)$.
- For $\boldsymbol{a} \in \{0,1\}^n$, define $v(\boldsymbol{a}) = \{i | a_i \neq 0, 1 \leq i \leq n\}$.

1. For $A \subseteq \{0,1\}^n$, define

$$H_A = \begin{bmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,n} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k,1} & h_{k,2} & \ldots & h_{k,n} \end{bmatrix},$$

   where $h_{i,j} \in GF(q)$ and H_A satisfies the following condition: The set $\{\boldsymbol{h}_{\cdot j} | j \in v(\boldsymbol{a} + \boldsymbol{a}')\}$, where $\boldsymbol{h}_{\cdot j}$ is the j-th column of H_A, is linearly independent over $GF(q)$ for any $\boldsymbol{a}, \boldsymbol{a}' \in A$.

2. An orthogonal design C for $A \subseteq \{0,1\}^n$ is defined by

$$C = \{\boldsymbol{x} | \boldsymbol{x} = \boldsymbol{r} H_A,\ \boldsymbol{r} \in GF(q)^k\}.$$

Orthogonal design: Example

Example

$q = 2, n = 4,\ A = \{0000, 1000, 0100, 0010, 0001, 1100, 1010, 1001\}$.

1. By the definition, the sets containing the following must be linearly independent:
   the 1st, 2nd, and 3rd columns of H_A;
   the 1st, 2nd, and 4th columns of H_A; and
   the 1st, 3rd and 4th columns of H_A.
   Then, the corresponding matrix H_A is given by
$$H_A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$
2. An orthogonal design C for A is defined by
$$\begin{aligned} C &= \{\boldsymbol{x} | \boldsymbol{x} = \boldsymbol{r} H_A,\ \boldsymbol{r} \in \{0,1\}^3\} \\ &= \{0000, 1000, 0101, 1101, 0011, 1011, 0110, 1110\}. \end{aligned}$$

Example: Posterior variance

Example: Posterior variance (n=2, q=3, N=9)

$$S_9 = \left[\left(\frac{9}{\sigma^2} (M^{-1})^* M^{-1} + S_0^{-1} \right)^{-1} \right],$$

where

$$(M^{-1})^* M^{-1} = \frac{1}{9} \begin{bmatrix} 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 2 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 4 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 4 \end{bmatrix}.$$