A Note on Bayesian Experimental Design Model Based on an Orthonormal System

Yoshifumi Ukita

Abstract

In this paper, analytical properties of Bayesian experimental design models based on an orthonormal system will be presented. The main idea of this paper is combining the models good for assuming a prior probability distribution over parameters with those good for deriving analytical properties. Firstly, it is shown that models expressed through the effect of each factor can be converted to those based on an orthonormal system. Next, it is shown that the posterior distribution and predictive distribution can be analytically derived in a Bayesian experimental framework. The result of this paper can be expected to be applied widely, especially in health care, where a Bayesian framework is necessary because the experiments are expensive.

Key Words:

Bayesian experimental design, machine learning, Fourier analysis

1 Introduction

Bayesian theory and methodology have seen dramatic growth in the last several decades. Lindley [1] reviewed experimental designs based on Bayesian decision theory. For sequential and non-sequential Bayesian methods based on a framework for optimal experimental designs, refer to [2, 3]. Bishop [4] introduced some analytical properties of Bayesian methods through models made from linear combinations of basis functions.

In experimental designs, since the traditional models are often expressed through the effect of each factor [5], these are adequate for assuming a prior probability distribution over the model parameters. However, these are not linear combinations of basis functions and are not good for deriving analytical properties within the Bayesian framework. In addition, not all parameters are independent, because there are constraints on the parameters.

Concerning the model, it was also shown that the model can be expressed in terms of orthonormal basis functions by using complex Fourier coefficients, making all parameters independent [6, 7, 8, 9].

In this paper, I will describe analytical properties of Bayesian experimental design models based on an orthonormal system. The main idea of this paper is combining the models good for assuming a prior probability distribution over parameters with those good for deriving analytical properties. Firstly it is shown that models expressed through the effect of each factor can be converted to those based on an orthonormal system. Using the results of [4], it is shown that the posterior distribution and predictive distribution can also be analytically derived within a Bayesian experimental framework. The result of this paper can be expected to be applied widely, especially in health care [10], where a Bayesian framework is necessary because the experiments are expensive.

This paper is organized as follows. In Section 2, I give the form of the single and multivariate Gaussian distribution necessary for this study as preliminaries. In Section 3, after providing notations for experimental designs, it is shown that models expressed through the effect of each factor can be converted to those based on an orthonormal system. In Section 4, it is shown that the posterior distribution and predictive distribution can also be analytically derived in a Bayesian experimental framework. Section 5 concludes this paper.

2 Preliminaries

2.1 Gaussian Distribution

In the case of a single variable x, the Gaussian distribution take the form

$$\begin{aligned}\mathcal{N}(x|\mu,\sigma^2)\\ =\ \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},\end{aligned} \tag{1}$$

where μ is the mean and σ^2 is the variance.

2.2 Multivariate Gaussian Distribution

In the case of an N-dimensional vector $\boldsymbol{x}$, the multivariate Gaussian distribution takes the form

$$\begin{aligned}&\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\\ &=\frac{1}{(2\pi)^{N/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},\end{aligned} \tag{2}$$

where $\boldsymbol{\mu}$ is an N-dimensional mean vector, $\boldsymbol{\Sigma}$ is an $N \times N$ covariance matrix, $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Sigma}^{-1}$ is the inverse of $\boldsymbol{\Sigma}$.

3 Experimental Design

In this section, after providing notations for experimental designs, I introduce two experimental design models.

Firstly, I explain the traditional model, which is expressed through the effect of each factor. This model clarifies how each factor affects the response variable [5]. Because I consider a Bayesian approach in this paper, I need to introduce a prior probability distribution over the model parameters. For the traditional model, it is easy to assume a prior probability distribution. However, this model is not good for deriving analytical properties of a Bayesian framework, because it is not linear combinations of basis functions. In addition, not all parameters are independent, because there are constraints on the parameters.

In contrast, models based on an orthonormal system are good for deriving analytical properties of Bayesian framework. However, it is not adequate to assume a prior probability distribution for models based on an orthonormal system, because the models are expressed by using complex Fourier coefficients.

Hence, if the former model can be converted to the latter, and vice versa, it is desirable in a Bayesian experimental design framework. I will show this in this section.

3.1 Notations for Experimental Designs

Let $F_1, F_2, \ldots, F_n$ denote the n factors to be included in an experiment. Suppose each factor has q levels, where q is a prime power.

Let the set $A \subseteq \{0,1\}^n$ denote all factors and interactions that might influence the response. For $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in A$, if $a_i = 0$ for all i $(0 \leq i \leq n)$, then $\boldsymbol{a}$ denotes the general mean. If $a_i = 1$ and $a_j = 0$ for all j $(j \neq i)$, then $\boldsymbol{a}$ denotes Factor F_i. If $a_i = 1$, $a_j = 1$ and $a_k = 0$ for all k $(k \neq i, j)$, then $\boldsymbol{a}$ denotes the interaction of Factor F_i and Factor F_j.

Let the set of index of factors $S_F = \{i | a_i = 1, \boldsymbol{a} \in A_1\}$, where $A_1 = \{\boldsymbol{a} | w(\boldsymbol{a}) = 1, \boldsymbol{a} \in A\}$ and $w(\boldsymbol{a})$ is the *Hamming weight* of $\boldsymbol{a}$. Let the set of index of interactions $S_I = \{\{i,j\} | a_i = 1, a_j = 1, \boldsymbol{a} \in A_2\}$, where $A_2 = \{\boldsymbol{a} | w(\boldsymbol{a}) = 2, \boldsymbol{a} \in A\}$.

3.2 Traditional Experimental Design Model

Let $t(\boldsymbol{x})$ denote the response of the experiment with level combination $\boldsymbol{x}$ and assume the model

$$t(\boldsymbol{x}) = \mu + \sum_{i \in S_F} \alpha_i(x_i) + \sum_{\{i,j\} \in S_I} \beta_{i,j}(x_i, x_j) + \epsilon, \tag{3}$$

where μ is the effect of general mean, $\alpha_i(x_i)$ is the effect of the x_ith level of

Factor F_i, $\beta_{i,j}(x_i, x_j)$ is the effect of the interaction of the x_ith level of Factor F_i and the x_jth level of Factor F_j, and ϵ is a zero-mean Gaussian random variable with variance σ^2. Here, generally, the constraints [5, p.249]

$$\sum_{\varphi=0}^{q-1} \alpha_i(\varphi) = 0, \tag{4}$$

$$\sum_{\varphi=0}^{q-1} \beta_{i,j}(\varphi, \psi) = 0, \tag{5}$$

$$\sum_{\psi=0}^{q-1} \beta_{i,j}(\varphi, \psi) = 0 \tag{6}$$

are assumed. Let the number of the independent parameters of (3) be K, and $\boldsymbol{u}$ denote a K-dimensional column vector.

Example 1 *Consider* $q = 3, n = 2$ *and* $A = \{00, 10, 10, 11\}$.

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

Using the constraints (4), (5) and (6), the independent parameters can be given as follows:
$\mu, \alpha_1(0), \alpha_1(1), \alpha_2(0), \alpha_2(1), \beta_{1,2}(0,0), \beta_{1,2}(0,1), \beta_{1,2}(1,0), \beta_{1,2}(1,1)$.
Hence, $\boldsymbol{u}$ *can be expressed by*

$$\boldsymbol{u} = \begin{bmatrix} \mu \\ \alpha_1(0) \\ \alpha_1(1) \\ \alpha_2(0) \\ \alpha_2(1) \\ \beta_{1,2}(0,0) \\ \beta_{1,2}(0,1) \\ \beta_{1,2}(1,0) \\ \beta_{1,2}(1,1) \end{bmatrix}, \tag{7}$$

□

3.3 Experimental Design Model Based on an Orthonormal System

Firstly, the levels of each factor can be represented by $GF(q)$, which is a Galois field of order q, and the level combinations can be represented by the n-tuples $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in GF(q)^n$. Then, the characters $\{\mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) | \boldsymbol{a} \in GF(q)^n\}$ form an orthonormal system. For a detailed information about characters, for example, refer to [11].

I use $t(\boldsymbol{x})$ to denote the response of the experiment with level combination $\boldsymbol{x}$ and assume the model [6]

$$t(\boldsymbol{x}) = \sum_{\boldsymbol{a} \in I_A} f_{\boldsymbol{a}} \mathcal{X}_{\boldsymbol{a}}(\boldsymbol{x}) + \epsilon, \tag{8}$$

where $I_A = \{(b_1 a_1, \ldots, b_n a_n) | \boldsymbol{a} \in A, b_i \in GF(q)\}$ and ϵ is a zero-mean Gaussian random variable with variance σ^2.

The parameters $\{f_{\boldsymbol{a}} | \boldsymbol{a} \in I_A\}$ are independent. Let $|I_A| = K$ and $I_A = \{\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_K}\}$.
Let $(f_{\boldsymbol{a_1}}, f_{\boldsymbol{a_2}}, \ldots, f_{\boldsymbol{a_K}})^T$ be denoted by $\boldsymbol{w}$.

Then, (8) can be also expressed by the equation

$$t(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{w} + \epsilon, \tag{9}$$

where

$$\boldsymbol{\phi}(\boldsymbol{x}) = \left[\mathcal{X}_{a_1}(\boldsymbol{x})\, \mathcal{X}_{a_2}(\boldsymbol{x}) \ldots \mathcal{X}_{a_K}(\boldsymbol{x})\right]^T.$$

Moreover, consider the relation between $\boldsymbol{u}$ in Sec. 3.2 and $\boldsymbol{w}$ in Sec. 3.3. In [8], the following equations about the relation are already provided.

$$\mu = f_{0\ldots0}. \tag{10}$$

$$\alpha_l(\varphi) = \sum_{\substack{a_l \in GF(q) \\ a_l \neq 0}} \mathcal{X}_{a_l}(\varphi) f_{0\ldots0a_l0\ldots0}. \tag{11}$$

$$\beta_{l,m}(\varphi, \psi) = \sum_{\substack{a_l \in GF(q) \\ a_l \neq 0}} \sum_{\substack{a_m \in GF(q) \\ a_m \neq 0}} \mathcal{X}_{a_l}(\varphi) \mathcal{X}_{a_m}(\psi) f_{0\ldots0a_l0\ldots0a_m0\ldots0}. \tag{12}$$

Using these equations, we can construct a $K \times K$ matrix $\boldsymbol{M}$ that satisfies the following equation:

$$\boldsymbol{u} = \boldsymbol{M}\boldsymbol{w}. \tag{13}$$

As the rank of $\boldsymbol{M}$ is apparently K, the inverse of $\boldsymbol{M}$ also exists. Hence, the following equation holds.

$$\boldsymbol{M}^{-1}\boldsymbol{u} = \boldsymbol{w}. \tag{14}$$

Using (13) and (14), the traditional model can be converted to the model based on orthonormal system, and vice versa. As we can see in the next section, this converting is desirable for deriving analytical properties in a Bayesian experimental design framework.

Example 2 *Consider* $q = 3$, $n = 2$ and $A = \{00, 10, 10, 11\}$. Then, $I_A = \{00, 10, 20, 01, 02, 11, 12, 21, 22\}$, *and* $\boldsymbol{w}$ *can be expressed by*

$$\boldsymbol{w} = \begin{bmatrix} f_{00} \\ f_{10} \\ f_{20} \\ f_{01} \\ f_{02} \\ f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \end{bmatrix}. \tag{15}$$

Let $\omega_3 = e^{2\pi i/3}$. *Using (10), (11), (12), and* $\mathcal{X}_l(k) = \omega_3^{lk}$ *from [11],* $\boldsymbol{M}$ *that satisfies (13) for (7) and (15) are given as*

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \omega_3 & \omega_3^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_3 & \omega_3^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & \omega_3 & \omega_3^2 & \omega_3 & \omega_3^2 \\ 0 & 0 & 0 & 0 & 0 & \omega_3 & \omega_3 & \omega_3^2 & \omega_3^2 \\ 0 & 0 & 0 & 0 & 0 & \omega_3^2 & 1 & 1 & \omega_3 \end{bmatrix}. \tag{16}$$

Next, I explain $\boldsymbol{M}^{-1}$. *First, let* r_1, s_1, s_2 *denote*

$$r_1 = \frac{1}{\omega_3^2 - \omega_3}, \tag{17}$$

$$s_1 = \frac{1 - \omega_3}{3(2 + \omega_3)}, \tag{18}$$

$$s_2 = \frac{1 + 2\omega_3}{3(2 + \omega_3)}. \tag{19}$$

Then, the following equations are holds.

$$\begin{bmatrix} r_1\omega_3^2 & -r_1 \\ -r_1\omega_3 & r_1 \end{bmatrix} \begin{bmatrix} \alpha_1(0) \\ \alpha_1(1) \end{bmatrix} = \begin{bmatrix} f_{10} \\ f_{20} \end{bmatrix}. \tag{20}$$

$$\begin{bmatrix} r_1\omega_3^2 & -r_1 \\ -r_1\omega_3 & r_1 \end{bmatrix} \begin{bmatrix} \alpha_2(0) \\ \alpha_2(1) \end{bmatrix} = \begin{bmatrix} f_{01} \\ f_{02} \end{bmatrix}. \tag{21}$$

$$\begin{bmatrix} s_1 & -s_2 & -s_2 & -1 \\ 1 & s_2 & s_1 & 1 \\ 1 & s_1 & s_2 & 1 \\ s_2 & -s_1 & -s_1 & -1 \end{bmatrix} \begin{bmatrix} \beta_{1,2}(0,0) \\ \beta_{1,2}(0,1) \\ \beta_{1,2}(1,0) \\ \beta_{1,2}(1,1) \end{bmatrix} = \begin{bmatrix} f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \end{bmatrix} \tag{22}$$

Hence, $\boldsymbol{M}^{-1}$ is given as follows.

$$\boldsymbol{M}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & r_1\omega_3^2 & -r_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -r_1\omega_3 & r_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r_1\omega_3^2 & -r_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -r_1\omega_3 & r_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & s_1 & -s_2 & -s_2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & s_2 & s_1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & s_1 & s_2 & 1 \\ 0 & 0 & 0 & 0 & 0 & s_2 & -s_1 & -s_1 & -1 \end{bmatrix}. \tag{23}$$

□

4 Bayesian Experimental Design Model Based on an Orthonormal System

In Sec.3, it is shown that models expressed through the effect of each factor can be converted to those based on an orthonormal system. As models based on an orthonormal system are linear combinations of basis functions, we can apply some analytical properties by Bishop [4] to this model. This section uses the results of [4, Section 3.3] to show that the posterior distribution and predictive distribution can also be analytically derived in the Bayesian experimental framework.

4.1 Likelihood Function

As explained in Sec. 3.3, it is assumed that target variable $t(\boldsymbol{x})$ is given by a deter-ministic function with additive noise, which is a zero-mean Gaussian random variable with variance σ^2 .

Hence, using (9) and (14), the likelihood function is given as

$$p(t|\boldsymbol{x}, \boldsymbol{u}, \sigma^2) = \mathcal{N}(t|\boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{M}^{-1}\boldsymbol{u}, \sigma^2) \tag{24}$$

Consider a data set of inputs $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ with corresponding target values $t(\boldsymbol{x}_1), \ldots, t(\boldsymbol{x}_N)$. Let the variables $\{t(\boldsymbol{x}_1), \ldots, t(\boldsymbol{x}_N)\}$ be a column vector denoted by $\mathbf{t}$.

Under the assumption that these target

values are drawn independently, the next expression for the likelihood function is obtained as

$$p(\mathbf{t}|\boldsymbol{X},\boldsymbol{u},\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(t(\boldsymbol{x}_n)|\boldsymbol{\phi}(\boldsymbol{x}_n)^T\boldsymbol{M}^{-1}\boldsymbol{u},\sigma^2). \quad (25)$$

Moreover, the likelihood function $p(\mathbf{t}|\boldsymbol{X},\boldsymbol{u},\sigma^2)$ can also be expressed by the equation

$$p(\mathbf{t}|\boldsymbol{X},\boldsymbol{u},\sigma^2) = \mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\boldsymbol{M}^{-1}\boldsymbol{u},\sigma^2\boldsymbol{I}), \quad (26)$$

where

$$\boldsymbol{\Phi}= \begin{bmatrix} \mathcal{X}_{a_1}(\boldsymbol{x}_1) & \mathcal{X}_{a_2}(\boldsymbol{x}_1) & \dots & \mathcal{X}_{a_K}(\boldsymbol{x}_1) \\ \mathcal{X}_{a_1}(\boldsymbol{x}_2) & \mathcal{X}_{a_2}(\boldsymbol{x}_2) & \dots & \mathcal{X}_{a_K}(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}_{a_1}(\boldsymbol{x}_N) & \mathcal{X}_{a_2}(\boldsymbol{x}_N) & \dots & \mathcal{X}_{a_K}(\boldsymbol{x}_N) \end{bmatrix}. \quad (27)$$

4.2 Bayesian Approach

In a Bayesian framework, we can assume a prior probability distribution of $\boldsymbol{u}$, which is denoted by $p(\boldsymbol{u})$. Under the observed data, $\boldsymbol{X}$ and $\mathbf{t}$, the posterior probability distribution of $\boldsymbol{u}$ can be calculated by using Bayes' theorem, which takes the form

$$p(\boldsymbol{u}|\boldsymbol{X},\mathbf{t}) = \frac{p(\mathbf{t}|\boldsymbol{X},\boldsymbol{u})p(\boldsymbol{u})}{p(\boldsymbol{X},\mathbf{t})}. \quad (28)$$

In other words, we can evaluate the uncertainty in $\boldsymbol{u}$ after we have observed $\boldsymbol{X}$ and $\mathbf{t}$, in the form of the posterior probability $p(\boldsymbol{u}|\boldsymbol{X},\mathbf{t})$.

4.3 Prior and Posterior Probability

As described, I assume a prior probability distribution over the model parameters $\boldsymbol{u}$. I will treat the variance σ^2 as a known constant.

Theorem 1 *Let the likelihood function be given by (26). The corresponding conjugate prior is given by a Gaussian distribution of the form*

$$p(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{u}|\boldsymbol{m}_0,\boldsymbol{S}_0). \quad (29)$$

Then the posterior probability is given by the equation

$$p(\boldsymbol{u}|\boldsymbol{X},\mathbf{t},\sigma^2) = \mathcal{N}(\boldsymbol{u}|\boldsymbol{m}_N,\boldsymbol{S}_N), \quad (30)$$

where

$$\boldsymbol{m}_N = \boldsymbol{S}_N\left(\frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^T\boldsymbol{\Phi}^T\mathbf{t} + \boldsymbol{S}_0^{-1}\boldsymbol{m}_0\right), \quad (31)$$

$$\boldsymbol{S}_N^{-1} = \frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{M}^{-1} + \boldsymbol{S}_0^{-1}. \quad (32)$$

Proof of Theorem 1:

Using Bayes' theorem, the posterior distribution can be written as

$$\begin{aligned} & p(\boldsymbol{u}|\boldsymbol{X},\mathbf{t},\sigma^2) \\ \propto\ & p(\mathbf{t}|\boldsymbol{X},\boldsymbol{u},\sigma^2)p(\boldsymbol{u}) \\ =\ & \mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\boldsymbol{M}^{-1}\boldsymbol{u},\sigma^2\boldsymbol{I})\mathcal{N}(\boldsymbol{u}|\boldsymbol{m}_0,\boldsymbol{S}_0) \\ \propto\ & \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{t}-\boldsymbol{\Phi}\boldsymbol{M}^{-1}\boldsymbol{u})^T(\mathbf{t}-\boldsymbol{\Phi}\boldsymbol{M}^{-1}\boldsymbol{u})\right\} \\ & \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{m}_0)^T\boldsymbol{S}_0^{-1}(\boldsymbol{u}-\boldsymbol{m}_0)\right\} \\ =\ & \exp\left\{-\frac{1}{2}\boldsymbol{u}^T\left(\frac{1}{\sigma^2}(\boldsymbol{M}^{-1})^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{M}^{-1} + \boldsymbol{S}_0^{-1}\right)\boldsymbol{u}\right. \end{aligned}$$

$$+\boldsymbol{u}^T \left(\frac{1}{\sigma^2} (\boldsymbol{M}^{-1})^T \boldsymbol{\Phi}^T \mathbf{t} + \boldsymbol{S}_0^{-1} \boldsymbol{m}_0 \right) + const \Bigg\}$$

$$= \exp\left\{ -\frac{1}{2} \left(\boldsymbol{u} - \boldsymbol{m}_N\right)^T \boldsymbol{S}_N^{-1} \left(\boldsymbol{u} - \boldsymbol{m}_N\right) + const \right\}, \tag{33}$$

where $\boldsymbol{m}_N$, $\boldsymbol{S}_N^{-1}$ *are given by (31), (32) respectively, and const denotes quantities independent of* $\boldsymbol{u}$. *We see that as a function of* $\boldsymbol{u}$, *this is also a quadratic form, and hence the posterior distribution will be Gaussian with mean vector* $\boldsymbol{m}_N$ *and covariance matrix* $\boldsymbol{S}_N$. □

Theorem 1 shows that the posterior distribution can also be analytically derived within a Bayesian experimental framework.

4.4 Predictive Distribution

In experimental designs, it is important to consider the predictive distribution. The predictive distribution [12] is defined by

$$p(t|\boldsymbol{x}, \mathbf{t}, \sigma^2) = \int p(t|\boldsymbol{x}, \boldsymbol{u}, \sigma^2) p(\boldsymbol{u}|\boldsymbol{X}, \mathbf{t}, \sigma^2) d\boldsymbol{u}. \tag{34}$$

The predictive distribution shown by Bishop [4, Sec. 3.3.2] can be also applied to this Bayesian experimental design framework.

If the conditional distribution $p(t|\boldsymbol{x}, \boldsymbol{u}, \sigma^2)$ is given by (24) and the posterior distribution is given by (30), then the predictive distribution is given as

$$p(t|\boldsymbol{x}, \mathbf{t}, \sigma^2) = \mathcal{N}(t|\boldsymbol{m}_N^T \boldsymbol{\phi}(\boldsymbol{x}), \sigma_N^2(\boldsymbol{x})), \tag{35}$$

where the variance $\sigma_N^2(\boldsymbol{x})$ of prediction distribution is given by

$$\sigma_N^2(\boldsymbol{x}) = \sigma^2 + \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{S}_N \boldsymbol{\phi}(\boldsymbol{x}). \tag{36}$$

Generally, it is difficult to derive the predictive distribution analytically. However, using experimental design models based on an orthonormal system, the predictive distribution can also be analytically derived in a Bayesian experimental framework.

5 Conclusion

In this paper, I described analytical properties of Bayesian experimental design models based on an orthonormal system. The main idea of this paper was combining the models good for assuming a prior probability distribution over parameters with those good for deriving analytical properties. Firstly, I showed that models expressed through the effect of each factor can be converted to those based on an orthonormal system. Next, using the results of [4], I showed that the posterior distribution and predictive distribution can also be analytically derived in a Bayesian experimental framework. The result of this paper can be expected to be applied widely, especially in health care, where a Bayesian framework is necessary because the experiments are expensive.

Acknowledgment

This research is partially supported by JSPS KAKENHI Grant Number JP17K00316 and the Academic Committee of Yokohama College of Commerce.

References

[1] D. V. Lindley, Bayesian Statistics, a Review, Philadelphia, 1972.

[2] J.O. Berger, Statistical Decision Theory, Foundations, Concepts and Methods, Springer, New York, 1980.

[3] K. Chaloner and I. Verdinelli, "Bayesian Experimental Design: A Review," Statistical Science, Vol.10, No.3, pp.273-304, 1995.

[4] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2010.

[5] H. Toutenburg and Shalabh, Statistical Analysis of Designed Experiments (Third Edition), Springer, 2009.

[6] Y. Ukita, T. Saito, T. Matsushima and S. Hirasawa, "A Note on the Relation between a Sampling Theorem for Functions over a $GF(q)^n$ Domain and Linear Codes," in Proc. 2009 IEEE Int. Conf. on SMC, pp.2665-2670, San Antonio, USA, Oct. 2009.

[7] 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.

[8] Y. Ukita and T. Matsushima, "A Note on Relation between the Fourier Coefficients and the Effects in the Experimental Design," in Proc. 8th Int. Conf. on Inf., Comm. and Signal Processing, Singapore, Dec. 2011.

[9] Y. Ukita, "Relation between the Fourier coefficients and the interaction effects expressed by using Kronecker product in the experimental design," The Yokohama Shodai Ronshu, Vol.49, No.1, pp.9-17, Sep. 2015.

[10] Design and Analysis of Experiments in Healthcare, Isaac Newton Institute for Mathematical Sciences, University of Cambridge, 6-10 July 2015.

[11] E.M. Stein, R. Shakarchi, Fourier Analysis: An Introduction, Princeton University

Press, 2003.

[12] T. Matsushima, H. Inazumi and S. Hirasawa, "An inductive inference procedure to minimize prediction error," IEEE Transactions on Systems, Man, and Cybernetics, Vol.23, No.2, pp.547-554, 1993.