Advances and Applications in Mathematical Sciences
Volume 20, Issue 9, July 2021, Pages 1937-1946

REDUCTION OF DIMENSIONALITY USING BAYESIAN APPROACH FOR SECOND ORDER RESPONSE SURFACE DESIGN MODEL

T. DEEPTHI, AMEEN SAHEB SHAIK
and N. CH. BHATRA CHARYULU

Bharath Institute of Engineering and Technology
Hyderabad-501504, India
E-mail: deepthitogercheti@gmail.com

Department of Science and Humanities
B. V. Raju Institute of Technology
Narsapur, Telangana, India
E-mail: ameensaheb.shaik@bvrit.ac.in

University College of Science
Osmania University
Hyderbad-7, India
E-mail: dwarakbhat@osmania.ac.in

Abstract

This paper provides the selection of best model for a second order response surface design in Bayesian approach. The posterior distribution of parameters obtained is presented and the method is illustrated with suitable examples under with and without restrictions on the moment matrix towards the orthogonality and rotatability. A comparison with classical methods is also presented.

1. Motivation

Let Y be the vector of response corresponding to a design matrix $X = ((x_{u_1}, x_{u_2}, \ldots, x_{u_v}))$, where x_{ui} be the level of the i^{th} factor in the u^{th} treatment combination. Assume the functional form of the response surface

2010 Mathematics Subject Classification: 62K20.
Keywords: Dimensionality reduction, Second order response surface design and Bayesian approach.
Received March 29, 2021; Accepted April 12, 2021

design model can be expressed as

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{1.1}$$

where $\underline{Y}_{N\times 1} = (Y_1, Y_2, \ldots, Y_N)'$ be the vector of observations, $X_{N\times p}$ be the Design matrix, $\underline{\beta}_{p\times 1}$ be the vector of parameters and $\underline{\varepsilon}_{N\times 1} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$ be the vector of random errors and assume that $\underline{\varepsilon} \sim N(0, \sigma^2 I)$.

The factor-response relationship is given by $E(Y) = f(x_1, x_2, \ldots, x_v)$ is called the 'Response Surface'. Design used for fitting the response surface models are termed as 'Response Surface Design'. The least square estimate of β is $\hat{\beta} = (X'X)^{-1} X'Y$ and the variance-covariance is $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$.

When the number of factors and order of model are increasing, the difficulty level of the analysis of the response surface design model increases. It leads to the problem of dimensionality of the model. If the number of factors is more, only a few factors are important and it is possible to eliminate the insignificant factors from the model, which are not affecting much the response. As a result, the time, cost, effort and data complexity can be minimized with the reduction of dimensionality of the model. Dimensionality reduction has enormous applications in various fields.

Only few authors made attempts on the reduction of size of the response surface design models. But, no much theoretical work is done on reduction of dimensionality in response surface designs and on its analysis.

Bayesian approach is a powerful technique can be used to estimate the uncertainties of parameters based on the posterior probabilities. Posteriors can be computed using the likelihood and prior distributions, since the marginal likelihood is simply a normalized constant, which need not be explicitly calculated.

Let $Y = [Y_1, Y_2, \ldots, Y_n]'$ be the observed sample drawn from a population with parameter θ. The joint density function of observed sample Y for the given parameter θ, called the likelihood function of Y denoted by $P(Y/\theta)$. Probability distribution for each parameter encapsulates the prior beliefs held about their most likely values called Prior distribution. $P(\theta)$ explains the information about the parameter θ called prior distribution of θ. An updated measure for each of the parameter values θ based on prior and given knowledge on Y called the posterior distribution of θ given Y.

Let $M_1, M_2, \ldots, M_m$ be the possible models $(m = 2^v)$ each containing subsets of $X_1, X_2, \ldots, X_v$. For each j, Model $M_j (j = 1, 2, \ldots, k)$ is defined by a family of distributions $P_{\theta j}$ where $P_{\theta j}$ has some prior distribution. The dimension of $P_{\theta j}$ for each j need not be same. The posterior probability for all possible models $M_1, M_2, \ldots, M_m$ can be evaluated based on their priors, likelihood and Normalized constant (unknown). Then, select the best model M^* with the highest posterior probability from $M_1, M_2, \ldots, M_m$.

2. Preliminaries (Concepts and Methods)

Consider the second order response surface design model

$$Y_u = \beta_0 + \sum_{i=1}^{v} \beta_i x_{ui} + \sum_{i=1}^{v} \beta_{ii} x_{ui^2} + \sum_{i<j}^{v} \beta_{ij} x_{ui} x_{uj} + \varepsilon. \tag{2.1}$$

It can be expressed in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_u \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1v} & x_{11}^2 & x_{12}^2 & \cdots & x_{1v}^2 & x_{11}x_{12} & \cdots & x_{1v-1}x_{1v} \\ 1 & x_{21} & x_{22} & \cdots & x_{2v} & x_{21}^2 & x_{22}^2 & \cdots & x_{2v}^2 & x_{21}x_{22} & \cdots & x_{2v-1}x_{2v} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{u1} & x_{u2} & \cdots & x_{uv} & x_{u1}^2 & x_{u2}^2 & \cdots & x_{uv}^2 & x_{u1}x_{u2} & \cdots & x_{uv-1}x_{uv} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nv} & x_{N1}^2 & x_{N2}^2 & \cdots & x_{Nv}^2 & x_{N1}x_{N2} & \cdots & x_{Nv-1}x_{Nv} \end{bmatrix}$$

$$\begin{bmatrix} \underline{\beta_0} \\ \beta_1 \\ \beta_2 \\ \vdots \\ \underline{\beta_v} \\ \beta_{11} \\ \beta_{22} \\ \vdots \\ \underline{\beta_{vv}} \\ \beta_{12} \\ \beta_{13} \\ \vdots \\ \beta v - 1v \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

$$Y = X\underline{\beta} + \underline{\varepsilon},$$

where, $Y = (Y_1, Y_2, Y_N)'$ is the vector of observations, $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_N)'$ is the vector of random errors and assume that $\underline{\varepsilon} \sim N(0, \sigma^2 I)$. $X_u = (1, x_{u1}, x_{u2} \ldots, x_{uv}, x_{u1}^2, x_{u2}^2 \ldots x_{uv}^2, x_{u1}x_{u2} \ldots x_{uv-1}x_{uv})$ is the u^{th} row of X $\underline{\beta} = (\beta_0, \beta_1, \beta_2 \ldots, \beta_v, \beta_{11}, \beta_{22} \ldots \beta_{vv}, \beta_{12} \ldots \beta_{v-1v})'$ is the vector of parameters. Let the total number of terms in the model be $k = (2v + vC_2)$.

The likelihood function of an observed sample Y with the given parameters β and σ^2 is

$$L(Y/\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-(2\sigma^2)^{-1} \sum_{i=1}^{N} (Y_i - X_i\beta)'(Y_i - X_i\beta)\right\}$$

$$L(Y, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (Y_i - X_i\beta)'(Y_i - X_i\beta)\right\}$$

$$\Rightarrow L(Y, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp$$

$$\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} [(Y_i - X_i\hat{\beta} + X_i\hat{\beta} - X_i\beta)'(Y_i - X_i\hat{\beta} + X_i\hat{\beta} - X_i\beta)]\right\}$$

$$\Rightarrow L(Y/\beta,\ \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp$$

$$\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[(Y_i - X_i\hat{\beta})'(Y_i - X_i\hat{\beta}) + (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})]\right\}$$

$$\Rightarrow L(Y/\beta,\ \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp$$

$$\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}[(Y_i - X_i\hat{\beta})'(Y_i - X_i\hat{\beta}) + (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})]\right\}$$

$$\frac{\partial L}{\partial \beta} \Rightarrow \hat{\beta} = (X'X)^{-1}X'Y \text{ and } \frac{\partial L}{\partial \sigma^2} \Rightarrow \hat{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta)}{N - k} \text{ where}$$

$$k = 2v + vC_2.$$

Then, the distribution of the parameters is:

$$\Rightarrow L(Y/\beta,\ \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp$$

$$\left\{-(2\sigma^2)^{-1}\sum_{i=1}^{N}[(Y_i - X_i\hat{\beta})'(Y_i - X_i\hat{\beta}) + (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})]\right\}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(Y - X\hat{\beta})^2 + \sum_{i=1}^{N} X_i^2(\beta - \hat{\beta})^2\right]\right\}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\hat{\sigma}^2(N - k) + \sum_{i=1}^{N} X_i^2(\beta - \hat{\beta})^2\right]\right\}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{\frac{-\hat{\sigma}^2(N - k)}{2\sigma^2}\right\} \exp\left\{\frac{1}{2\sigma^2/\Sigma X_i^2}(\beta - \hat{\beta})^2\right\}$$

$$= (\sqrt{2\pi})\sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2/\Sigma X_i^2}(\beta - \hat{\beta})^2\right\} \cdot (\sigma^2)^{-(N-1)/2}$$

$$\cdot \exp\left\{\frac{-\hat{\sigma}^2(N - k)}{2\sigma^2}\right\}.$$

Then the posterior distribution of the parameter, $\hat{\beta}$ is

$$P(\hat{\beta}/X, Y) \propto \sigma^{-N-1} \exp$$

$$\left\{-\frac{1}{2\sigma^2}[(Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})]\right\}$$

$$\propto \sigma^{-N-1} \exp\left\{-\frac{1}{2\sigma^2}[(N-k)\hat{\sigma}^2 (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})]\right\}$$

since

$$(N-k)\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$\propto \sigma^{-N-1} \exp\left\{-\frac{1}{2\sigma^2}(\beta - \hat{\beta})' X'X(\beta - \hat{\beta})\right\}$$

$$\sim MVN\,(\hat{\beta}, (X'X)^{-1}\sigma^2). \tag{2.2}$$

And, the posterior distribution of the parameter, $\hat{\sigma}^2$ is

$$P(\hat{\sigma}^2/X, Y) \propto \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2}\{(n-k)\hat{\sigma}^2 + (\beta - \hat{\beta})X'X(\beta - \hat{\beta})\}\right)$$

$$\propto \sigma^{-(n-k+1)} \exp\left(-\frac{1}{2\sigma^2}(N-k)\hat{\sigma}^2\right)$$

$$\sim IG\left(\frac{(N-k)}{2}, \frac{(N-k)\hat{\sigma}^2}{2}\right)$$

where

$$k = 2v + vC_2. \tag{2.3}$$

The method for reduction of second order response surface design model in case of without restrictions and restrictions towards rotatability on moment matrix are illustrated in the examples 2.1 and 2.2 are presented below.

Example 2.1. Consider the experimental data with four factors and sixteen design points:

Let Y = [53.3, 78, 62.4, 78.9, 75.9, 75.4, 71.3, 84.4, 64.5, 67.5, 72.8, 85.3, 71.4, 83.3, 82.9, 81.7]′ be the vector of responses obtained at 16 design points [(-1, -1, -1, -1), (1, -1, -1, -1), (-1, 1, -1, -1), (1, 1, -1, -1), (-1, -1, 1, -1), (1, -1, 1, -1), (-1, 1, 1, -1), (1, 1, 1, -1), (-1, -1, -1, 1), (1, -1, -1, 1), (-1, 1, -1, 1), (1, 1, -1, 1), (-1, -1, 1, 1), (1, -1, 1, 1), (-1, 1, 1, 1), (1, 1, 1, 1)]′ respectively.

Using R the posteriors for all the 16,384 possible linear combination models are evaluated and the Bayesian estimated parameter values are: $\hat{\beta}_0 = 74.31$; $\hat{\beta}_1 = 5$; $\hat{\beta}_2 = 3.149$; $\hat{\beta}_3 = 3.974$; $\hat{\beta}_4 = 1.86$; $\hat{\beta}_{12} = 0.112$; $\hat{\beta}_{13} = -2.089$; $\hat{\beta}_{14} = -1.725$; $\hat{\beta}_{23} = -1.366$; $\hat{\beta}_{24} = 1.349$; $\hat{\beta}_{34} = 0.326$.

The resulting reduced Bayesian model is $Y = 74.312 + 5X_1 + 3.15X_2 + 3.975X_3 + 1.862X_4 - 2.088X_{13} - 1.725X_{14}$, with error sum of squares 231.493 with 9 degrees of freedom and with an R^2 value is 0.810 selected with highest probability value 0.025 with six variables.

Example 2.2. A chemical Engineer is investigating the yield (Y) of a process. Three process variables are of interest: Temperature (A), Pressure (B), and Catalyst Concentration (C). Each variable can be run at a low and a high level, and the engineer decides to run a 2^3 design with four center points. The design and the resulting yields are shown below.

Let Y = [32 46 57 65 36 48 57 50 44 53 56]’ be the yield of a process corresponding to coded variables X_1, X_2 and X_3 are [(-1 -1 -1), (1 -1 -1), (-1 1 -1), (1 1 -1), (-1 -1 1), (1 -1 1), (-1 1 1), (0 0 0), (0 0 0), (0 0 0), (0 0 0)]’ respectively. Using R the posteriors for all the 512 possible linear combination models are evaluated and the Bayesian estimated parameter values are: $\hat{\beta}_0 = 10$; $\hat{\beta}_3 = 0.499$; $\hat{\beta}_{33} = -0.249$; $\hat{\beta}_{12} = -1.50$; $\hat{\beta}_{13} = -0.501$; $\hat{\beta}_{23} = -0.999$.

The resulting reduced Bayesian model is $Y = 50.906 + 5.484X_1 + 10.484X_2$, with error sum of squares 98.906 with 8 degrees of freedom and with an R^2 value is 0.895 selected with highest probability value 0.128 with two variables.

3. Results

The reduced models using different approaches like Bayesian, Nested, Stepwise, Forward, Backward elimination and an analysis on reduced Second Order Response Surface Design Model in case of orthogonal and non orthogonal are presented with Sum of Squares due to residual, Significance value, R^2 value are evaluated.

Table 3.1. Comparison of Models for Data in example 2.1.

S.N.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	74.313	74.313	74.313	74.313	74.313	74.313
2	β_1	5.000	5	5	5	5	5
3	β_2	3.15	–	–	3.15	3.15	3.15
4	β_3	3.975	3.975	3.975	3.975	3.975	3.975
5	β_4	1.862	–	–	–	–	1.862
6	β_{12}	0.113	–	–	–	–	–
7	β_{13}	– 2.088	–	–	–	–	– 2.088
8	β_{14}	– 1.725	–	–	–	–	– 1.725
9	β_{23}	– 1.363	–	–	–	–	–
10	β_{24}	1.35	–	–	–	–	–
11	β_{34}	– 0.325	–	–	–	–	–
SSR		1045.16	652.81	652.81	811.57	811.57	984.405
ESS		170.738	563.088	563.088	404.328	404.328	231.493
MSE		34.148	43.314	43.314	33.694	33.694	25.721
R^2 value		0.860	0.537	0.537	0.667	0.667	0.810

Note. The reduced model in all the approaches is not same and the mean square error values for full and Bayesian reduced models differ. It can be noted that the values of parameters β_4, β_{11}, β_{22}, β_{33}, β_{44}, β_{12}, β_{13}, β_{14},

$\beta_{23}, \beta_{24}, \beta_{34}$ are insignificant in nested and backward approaches whereas in Bayesian $\beta_{11}, \beta_{22}, \beta_{33}, \beta_{44}, \beta_{12}, \beta_{23}, \beta_4, \beta_{34}$ are insignificant and in stepwise and forward $\beta_2, \beta_4, \beta_{11}, \beta_{22}, \beta_{33}, \beta_{44}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{23}, \beta_{24}, \beta_{34}$ are insignificant. Cumulative Posterior Probability value for best five models is 0.1051.

Table 3.2. Comparison of Models for Data in example 2.2.

S.No.	Parameters	Full Model	Stepwise	Forward	Backward	Nested	Bayesian
1	β_0	50.75	50.906	50.906	50.906	50.815	50.906
2	β_1	5	5.484	5.484	5.484	5.348	5.484
3	β_2	10	10.484	10.484	10.484	9.618	10.484
4	β_3	0.5	–	–	–	–	–
5	β_{33}	– 0.25	–	–	–	–	–
6	β_{12}	– 1.5	–	–	–	–	–
7	β_{13}	– 0.5	–	–	–	–	–
8	β_{23}	– 1	–	–	–	–	–
SSR		861.977	841.821	841.821	841.821	841.821	841.821
ESS		78.75	98.906	98.906	98.906	98.906	98.906
MSE		26.25	12.363	12.363	12.363	12.363	12.363
R^2 value		0.916	0.895	0.895	0.895	0.895	0.895

Note. The reduced models in all the approaches is same and the mean square due to residual and R^2 values are same. Cumulative Posterior Probability value for best five models is 0.3502.

4. Remarks

Bayesian approach reduces the size of the original model by selecting the significant factors from the original set of variables based on the posterior probability values.

In Bayesian, manual computation is difficult and time consuming but provides better or equally efficient reduced models when compared with other traditional approaches like stepwise, Forward, Backward elimination and Nested.

Bayesian methods do not require an analytical likelihood function whereas the maximum likelihood estimator requires analytical likelihood function and behavior of likelihood function are crucial.

Bayesian method is complicated when compared to least squares method because Bayesian procedure depends on distribution for generation of large samples, which can be used to compute the parameters of interest using Gibbs Sampling algorithm.

References

[1] A. K. Bansal, Bayesian Parametric Inference, Narosa Publishing House, New Delhi, 2007.

[2] N. Ch. Bhatra Charyulu and S. K. Ameen Saheb, Note on Reduction of Dimensionality for Second Order Response Surface Design Model, Communications in Statistics-Theory and Methods, Taylor and Francis series 46 (2017), 3520-3525.

[3] A. I. Khuri and J. A. Cornell, Response surface, Design and Analysis, 2nd edition Marcel Dekker, Newyork, 1996.

[4] T. J. Mitchell and J. J. Beauchamp, Bayesian Variable selection the Linear Regression, Journal of the American Statistics Association 83 (1988), 1023-1036.

[5] R. B. O'Hara and M. J. Sillanpaa, A Review of Bayesian Variable Selection Methods: What, How and Which, Bayesian Analysis 4(1) (2009), 85-118.

[6] T. Deepthi S. K. Ameen Saheb and N. Ch. Bhatra Charyulu, Dimensionality Reduction in First order Response Surface Design Model Using Bayesian Approach, Far East Journal of Theoretical Statistics 54(4) (2018), 327-340, ISSN No 0972-0863.