BAYESIAN APPROACH TO THE D-OPTIMAL FOR MIXTURE EXPERIMENTAL DESIGN

UQWATUL ALMA WIZSA[1*], UTAMI DYAH SYAFITRI[1], AJI HAMIM WIGENA[1].

[1]Department of Statistics, IPB University, INA

*uqwatulalma@gmail.com

ABSTRACT

A mixture experiment is a special case of response surface methodology in which the value of the components is proportions. In case there are constraints on the proportions, the experimental region can be not a simplex. The classical designs such as a simplex-lattice design or a simplex-centroid design, in some cases, cannot fit the problem. In this case, the optimal design comes up as a solution. A D-optimal design is seeking a design in which minimizing the covariance of the model parameter. Some model parameters are important and some of them are less important. As the priority of the parameters, the prior information of parameters is needed in advance. This brings to a Bayesian D-optimal design. The focus in this research was on a baking experiment in which consisted of three ingredients with lower bounds on the proportion of the ingredients. The assumption model was a quadratic model. Due to the priority of the model parameters, the Bayesian D-optimal design was used to solve the problem. A point-exchange algorithm was developed to find the optimal design. Nineteen candidates were used to choose twelve design points. It found that the potential term is feasible to the actual model and design points represent overall points in the design area.

Key Words: Bayesian, D-optimal design, mixture experiment, a point-exchange algorithm

1 Introduction

Mixture experimental design used especially in the industrial sector. A Mixture experiment is a design in which the components is proportions. The proportions lie between 0 and 1 and the sum of the proportions among the components is unity [1]. Due to the restriction of the mixture experiment, the proportions of the components are dependence [2].

Unlike other designs, the compositions of a mixture design depend on the assumption model. The model parameters are some important but others are not. The prior information on the parameters is needed. Besides, there are some constraints on the proportions. The constraints affect the experimental region. In some cases, the classical mixture designs cannot fit the problem. Hence, it needs an optimal design approach. The optimal design is seeking a design based on a certain criterion. A D-optimality criterion is a criterion that widely used in mixture experiments. It can reduce the uncertainty of estimating the parameters model [3]. This criterion is useful for the performance of design under the assumption of the model [4]. The "D" stands for determinant. The D-optimality criterion is a criterion that minimizing the determinant of the inverse of the information matrix or it is equivalence with maximizing the determinant of the information matrix [5].

The optimal design based on the matrix of design very depends on the assumption of the model [6]. Inaccuracies in model assumption impact on inaccuracies in design result. The D-

optimality criterion with the Bayesian approach can retrieve basic information from the distribution of the parameter model, so the dependence of D-optimal to model that assumed can be reduced.

2   Theory

2.1   Bayesian D-optimal Design

The model assumed used the quadratic canonical Shefee model [7]. The general form of the quadratic canonical model is

$$\eta(x) = \sum_i \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j \tag{1}$$

The total proportion of components which amounted to 100% in the mixture caused the linear model do not have intercept. Then, the cross product $x_i x_j$ and x_i^2 cannot be included together to the model, because it caused perfect collinearity [8]. Estimating model parameters using least squares estimation can be calculated as

$$b = (X^T X)^{-1} X^T Y \tag{2}$$

The variance of β is $\sigma^2 (X^T X)^{-1}$. Minimizing the variety of β can be calculated as a maximize of the determinant $(X^T X)$.

In the Bayesian approach of D-optimal design, matrix X divided into two conditions, namely primary and potential terms. The most important term that wants to fit in the model called primary terms and the term that potential (possibly important) to the model called potential terms [9]. Bayesian used initial information to predict the condition. Thereafter, the initial information is taken from estimating parameters of primary and potential terms. The initial information can be formed into prior distribution. The prior distribution of the primary term is $\beta_{pri} \sim N(\mu, \sigma^2)$ and the prior distribution of the potential term is $\beta_{pot} \sim N(0, \tau^2 \sigma^2 I)$. The value of τ is the ratio of the variety on the potential terms with the error rate. A large value of τ means that some of the potential terms are feasible on the model. Then, a small value of τ means all potential terms should not be included in the model [6].

Based on *prior distributions* above, *posterior distribution* to parameters of model [9,10] is

$$p(\beta|y, \sigma) \sim N \left[(X^T X + \frac{K}{\tau^2})^{-1} X^T Y, \sigma^2 (X^T X + \frac{K}{\tau^2})^{-1} \right] \tag{3}$$

K is $(p + q) \times (p + q)$ identity matrix with p diagonal is zero and q others diagonal is one. The variance of the posterior distribution is $\sigma^2 (X^T X + K/\tau^2)^{-1}$. So, minimize the variance similar to maximize the value of the determinant information matrix below [11]

$$(X^T X + K/\tau^2) \tag{4}$$

DuMouchel and Jones also introduce the procedure called scaling convention [9]. This procedure is to make the effect of the potential term can be interpreted. Then, the scaling convention also minimizes the correlation between primary and potential terms. Let $X = [X_{pri}|X_{pot}]$ is a design matrix that divided into primary and potential terms. Each unconstant primary term transformed by interval -1 to 1. Then potential terms transformed such that max(X_{pot})-min(X_{pot})=1 [9].

For each primary term, the midpoint value is $M = (L + U)/2$ and the half of the range is $\Delta = (U - L)/2$, L and U is a lower and an upper value of the primary term. Then the scaled, l_k, from the primary term x_k is

$$x_k = (l_k - M)/\Delta \tag{5}$$

Potential terms regressed to primary terms, by least square estimation the coefficients regression (α) of X_{pot} to X_{pri} is

$$\alpha = (X_{pri}^T X_{pri})^{-1} X_{pri}^T X_{pot} \tag{6}$$

Defined R residual of regression of X_{pot} to X_{pri}, $R = X_{pot} - X_{pri}\alpha$ and Z is the transformation of R.

$$Z = R/(max(R) - min(R)) \tag{7}$$

The result, definition of $X = [X_{pri}|X_{pot}]$ becomes $X = [X_{pri}|Z]$.

2.2 Point Exchange Algorithm

The process of finding optimal design points from candidate points is carried out with an algorithm. One simple algorithm that can be used is the point exchange algorithm. This algorithm starts by randomly selecting the design points on the set of candidate points as many as n points as the initial design. Next, one by one the points in the initial design are exchanged with one other point from the set of candidate points in sequence. This process is carried out to improve the initial draft criteria that have been selected.

The randomization of points for the initial design does not guarantee that the design obtained is the optimal design. The point replacement process also does not find many permutations from all of the candidate points owned. Therefore, some initial designs were chosen randomly and carried out iteratively to overcome the limitations of point changes.

3   A Practical Example

As a practical use of the D-optimal design using the Bayesian approach, an example of a mixture experimental design with three components and constraints function was presented. Defined the first component as x_1, the second component as x_2, and the third component as x_1. The constraint function of the first and second components greater than or equal to 0.1 and third component greater than or equal to 0.6.

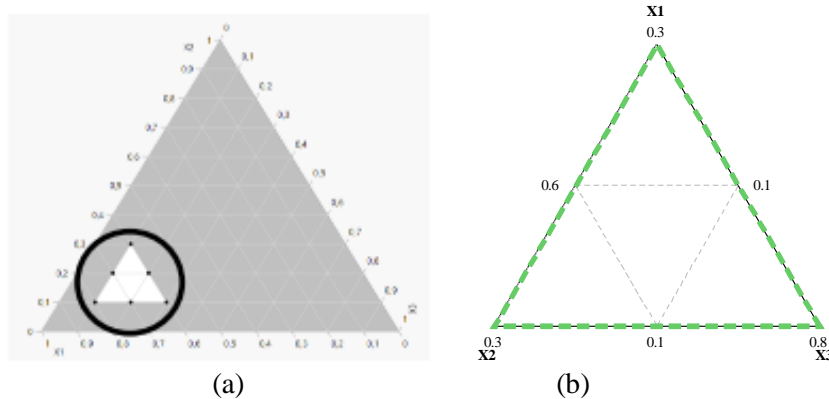(a)   (b)

Figure 1: Design region in the full area (a) Design region by the constraint function (b)

The constraint function of each component can figure out as a design region. The region with constraints function is a small part of the full design area (Figure 1a). If the region was magnified, it can be seen as Figure 1b and points along the design can be chosen as the candidate point. Points on the edge of the region called extreme vertices. It represents the maximum proportion of each component. The other points can be in the middle of the line, within, or center of the region.

Nineteen candidates generated using Cox-direction. Define the first component as x_i and other components as x_j, so changes for every component can be calculated $\tilde{x}_i = x_i + \delta_i$ dan $\tilde{x}_j = x_j - \delta_i x_j /(1 - x_i)$ for $i \neq j$ and $i = 1,2, \dots, p$, p is the number of candidates that can be generated [12]. A list of candidates in this case, can be shown in table 1.

Table 1: Candidates Set of Mixture

No	Proportion		
	x_1	x_2	x_3
1	0.1000	0.1000	0.8000
2	0.1000	0.1500	0.7500
3	0.1000	0.2000	0.7000
4	0.1000	0.2500	0.6500
5	0.1000	0.3000	0.6000
6	0.1333	0.1333	0.7334
7	0.1333	0.2333	0.6334
8	0.1500	0.1000	0.7500
9	0.1500	0.1500	0.7000
10	0.1500	0.2000	0.6500
11	0.1500	0.2500	0.6000
12	0.1666	0.1667	0.6667
13	0.2000	0.1000	0.7000
14	0.2000	0.1500	0.6500
15	0.2000	0.2000	0.6000
16	0.2333	0.1333	0.6334
17	0.2500	0.1000	0.6500
18	0.2500	0.1500	0.6000
19	0.3000	0.1000	0.6000

The candidates of the design above can be figured in the design region as figure 2. Nineteen candidates chose based on the great position of the design in the region. So the candidates can represent overall of the design.
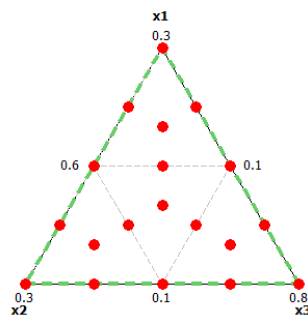

Figure 2: Design region of candidates design

The model used the canonical quadratic model, it was written as

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \beta_6 x_{2i} x_{3i} \tag{8}$$

which

y_i = response, $i = 1,2,...,n$.

β_j = model parameters, $j = 1,2,...,6$.

x_{ji} = value of independent variables $i = 1,2,...,n$, $j = 1,2,...,6$.

From the model above, the design matrix can be constructed as

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{11}x_{12} & x_{11}x_{13} & x_{12}x_{13} \\ x_{21} & x_{22} & x_{23} & x_{21}x_{22} & x_{21}x_{23} & x_{22}x_{23} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n1} & x_{n3} & x_{n1}x_{n2} & x_{n1}x_{n3} & x_{n2}x_{n3} \end{bmatrix} \tag{9}$$

The first-order of model assumed as primary terms, and the second-order assumed as potential terms. Then to calculate the determinant of the matrix information K matrix can be used as

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

From the design matrix, the matrix information can be constructed to get an optimal design from several values of τ by equation 4. The determinan of matrix information from the optimal design of several τ describe in Figure 3.
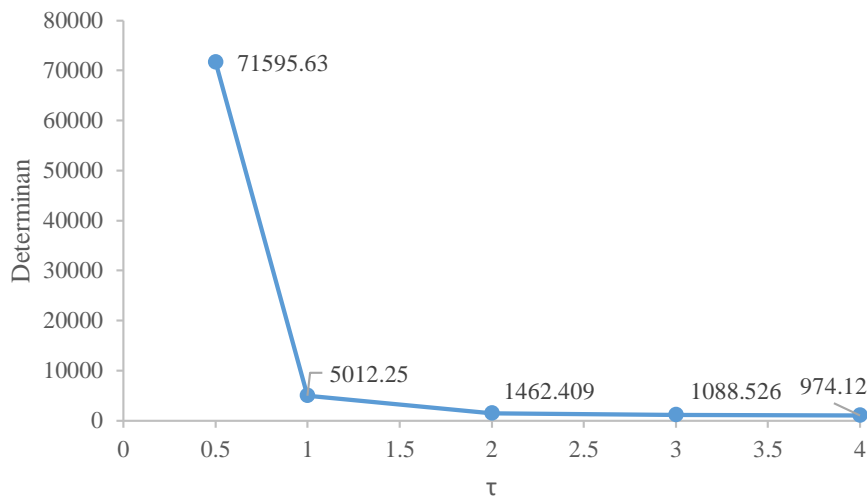

Figure 3: The determinant of several values of τ

Figure 3 described the determinant of designs in difference of τ. In $\tau = 0.5$ the determinant is 71595.63, then significantly decrease in $\tau = 1$ that the determinant is 5012.25. From $\tau = 2$, $\tau = 3$, and $\tau = 4$ the determinant have not many changed, the determinant 1462.409, 10808.529, 974.12 respectively. So, by the exploration in determinant, the design can be optimal in $\tau = 2$, because the determinant start to convergent there.

The result of design points also can form as the design region below.
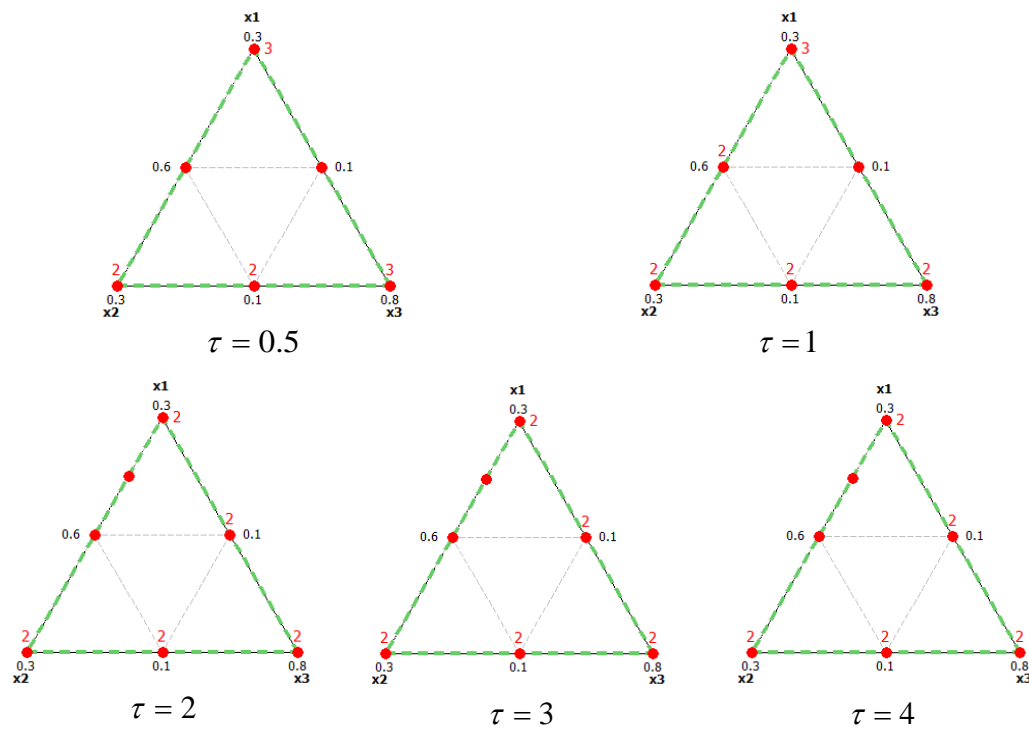
Figure 4: Design region for several value of τ

The design in $\tau = 0.5$ have no different with design in $\tau = 1$. The design point detects in every extreme point and half of the design region. But, some points of design have different numbers of replication in both designs. It means that the design could be changed for higher τ. If the value of τ increase, when $\tau = 2$ the design points were similar to $\tau = 3$ and $\tau = 4$. Not only in the points of the design but also in the number of replication for all points. The design has added one point from the design in $\tau = 1$. So, from this result, the design constant start from $\tau = 2$, and this design was chosen as optimal design points. The result of optimal design points using Bayesian D-optimal design in two components of the mixture can be seen in table 2. The determinant for this design about 1462.409 with seven different points of design.

Table 2: Design points of optimal design

No	Proportion		
	x_1	x_2	x_3
1	0.20	0.10	0.70
2	0.10	0.20	0.70
3	0.10	0.20	0.70
4	0.10	0.30	0.60
5	0.30	0.10	0.60
6	0.25	0.15	0.60
7	0.10	0.10	0.80
8	0.20	0.20	0.60
9	0.10	0.30	0.60
10	0.30	0.10	0.60
11	0.20	0.10	0.70
12	0.10	0.10	0.80
Determinant	1462.409		

Table 2 represents the proportion of each component of the mixtures as the result of optimal design by Bayesian D-optimal design. There are twelve points with seven different

points. This combination of points can be running in the laboratory as the optimal composition of the component in the mixture.

4   Conclusion

Bayesian D-optimal design can be used as an alternative to constructing the optimal design of the mixture. By using the D-optimal criterion approach to the Bayesian the composition of components in mixture will not depend on the model that assumed. From the practical example in three constraint components in a mixture, seven different points from twelve points were constructed. The design convergent in $\tau = 2$ with determinant about 1462.409.

5   Acknowledgments

Reference

[1]   W.F. Smith, *Experimental Design for Formulation*, Virginia American Statistical Association and the Society for Industrial and Applied Mathematics, 2005.

[2]   J. A. Cornell, *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data, Third Edition,* New York (US): John Wiley & Sons, Inc, 2002.

[3]   T.J. Mitchell, "An Algorithm for the Construction of "D-optimal" Experimental Designs", *Technometrics*, vol. 16, no. 2, pp. 203-210, 1974.

[4]   A. DasGupta, "Review of Optimal Bayes Designs", *Handbook of Statistics*, *Elsevier,* vol. 13, pp 1090-1147, 1996.

[5]   K. Chaloner and I. Verdinelli, "Bayesian Experimental Design: A Review", *Statist. Sci 10,* no. 3, pp. 237-304. 1995.

[6]   L. Hefang, R.H. Myers, and Y. Keying. "Bayesian Two-Stage Optimal Design for Mixture Models", *Journal of Statistical Computation and Simulation*, vol. 66, no. 3, pp. 209-231, 2000.

[7]   H. Schefee, "Experiments with Mixtures", *Journal of the Royal Statistical Society, B* 20, pp 344-360, 1958.

[8]   A. Ruseckaite, P. Goos, D. Fok, "Bayesian D-optimal Choice Design for Mixtures", Rotterdam: *Journal of the Royal Statistical Society: Series C Applied Statistics*, vol. 66, 2016.

[9]   W. DuMouchel and B. Jones. "A Simple Bayesian Modification of D-optimal Design to Reduce Dependence on an Assumed Model", *Technometrics,* vol. 36, no. 1, pp. 37-47, 1994.

[10]   J. Pilz, *Bayesian Estimation and Experimental Design in Linear Regression Models*, New York: John Wiley & Sons, Inc, 1991.

[11]   J. Andere-Rendon, D.C. Montgomery, and D.A. Rollier, "Design of Mixture Experiments Using Bayesian D-Optimality", vol. 29, no. 4, pp. 451-463, *Journal of Quality Technology*, 1997.

[12]   W.D. Rahayu, *Rancangan D-optimal Untuk Respon Ordinal Pada Rancangan Campuran,* Bogor: Institut Pertanian Bogor, 2019.