D-optimal design for ordinal responses in mixture experiments

To cite this article: W D Rahayu *et al* 2019 *J. Phys.: Conf. Ser.* **1317** 012017

View the article online for updates and enhancements.

D-optimal design for ordinal responses in mixture experiments

W D Rahayu*, U D Syafitri, A M Soleh

Department of Statistics, Bogor Agricultural University, Bogor, Indonesia
Level 4 Wing 22 Meranti Street, Dramaga, 16680

*wikasantidr022@gmail.com

Abstract. A mixture experiment is an experiment in which the response depends on the ingredients proportions. In organoleptic test, the response is ordinal. This research focused on how to generate the D-optimal design of ordinal response. The model belongs to Generalized Linear Model (GLM) with logit as the link function. To find the D-optimal design which is seeking the design that minimizing the covariance of parameter model, the point-exchange algorithm was constructed. There were three ingredients of a cookies recipe which were needed to mix meanwhile other ingredients were fixed. The D-optimal design of the three ingredients consisted of seven design points.

1. Introduction
Experimental designs are applied in an industry in order to find a final good product. The experimental design is an efficient way to have a quality product compared to the classical way: trial and error. This is because in the design of experiments, there is a systematic procedure or planning and also a series of testing in order to have a good final product. One of experimental design which is widely used in industry to formulate a product is a mixture experiment. A mixture experiment is an experiment in which the response depends on the ingredients proportions not the total amount. The mixture experiments are implemented in various industries such as food industry, pharmacy industry, etc.

In the food industry, one of the evaluation of the product is its taste. To evaluate the product, the likert scale is used. In statistics, the likert scale belongs to ordinal scale. As the likert scale is a response, hence the problem come up. Existing mixture designs are available in commercial software is based on linear model in which the response is numerical scale. Mancenido (2016) proved that if the response was ordinal and the design was based on linear model, the goodness fit of models is only 35%. Therefore, it is necessary to construct an optimal design based on an ordinal response.

Optimal designs are a branch of experimental design in which has flexibility to construct a noptimal design based on a real situation. The researchers make a design based on a certain optimality criterion (Goos, 2002). The criterion that widely used is D-optimality criterion which is focused on precision of parameter. In terms of ordinal response, a D-optimal design is seeking a design which has minimum of a generalized variance of the estimated regression coefficients. To find the optimal design based on the real condition, an algorithm is needed. As the final design is found by an algorithm, so the design could be an optimal or an nearly optimal design.

This research in this paper focused on generating the optimal mixture design when the response was an ordinal response. The model used was a proportional-odds model with a logistic link. The model

belongs to Generalized Linear Models (GLM) because the response follows non-normal distribution (ordinal logistic distribution). Furthermore, a point-exchange was developed to find the D-optimal design for ordinal response. The algorithm was implemented on the recipe of cookies in which involved three main ingredients.

2. Method

The cookies recipe tried to mix three flour meanwhile other ingredients were fixed. The tree flour were flour 1 (x_1), flour 2 (x_2), and flour 3 (x_3). Each flour had different constraints and it is shown in Table 1. Flour 1 and flour 2 are at least 10% in a dough, respectively. In contrast, flour 3 is at least 60% in a dough.

Table 1 The constraints of each ingredients

Component	Ingredients
1	$x_1 \geq 0.1$
2	$x_2 \geq 0.1$
3	$x_3 \geq 0.6$

Based on the constraints, the experimental region of the mixture experiment is shown in Figure 1. The experimental region is an triangular region but a part of a whole simplex.
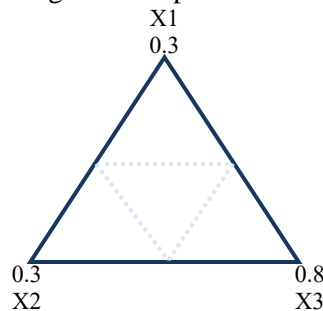

Figure 1 The experimental region of the mixture problem

To construct the design, a point-exchange algorithm was build. Steps of the point-exchange algorithm to construct the optimal design are shown below:

Simulation of parameter selection

Determine the assumption of the model.

For an ordinal response, the proposional odds form of the logits of cumulative probabilities as shown as:

$$logit\ P(y \leq j) = \log\frac{P(y \leq j)}{P(y > j)} = \alpha_j + \boldsymbol{\beta}'\boldsymbol{x}\ ;\ \alpha_1 < \alpha_2 < \cdots < \alpha_{c-1}$$

where:

y : response variable, having $c = 1,2,\ldots,j$ ordered categories

α_j : unique effects for each *logit* or category

$\boldsymbol{\beta}$: covariate effects

$\boldsymbol{x}$: explanatory variable

For a c-category response, there are $c - 1$ logits, one for each level, with the exception of the lat category. In this case, the model used was a quadratic model, the parameterization of the proportional odds model with three response categories and three components with linear model terms formed as:

$$\eta_{i1} = \alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3$$
$$\eta_{i2} = \alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3$$

Determine the initial design used an extreme vertex design with the D-Optimal criterion.

The number of the experimental runs n = 150. The design matrix can be written as:

$$X = \begin{bmatrix} 0 & x'_1 \\ 1 & x'_1 \\ \vdots & \vdots \\ 1 & x'_1 \\ \vdots & \vdots \\ 0 & x'_{150} \\ 1 & x'_{150} \\ \vdots & \vdots \\ 1 & x'_{150} \end{bmatrix}$$

Determine the initial of the parameters.

The parameter model, **β**, was assumed to follow uniform distribution. In this case, **β** ~ U (a,b). The range of parameter estimates is shown in table 2.

Table 2 The range of parameter estimation

Parameter	Distribution
β_1	$U \sim (3, 5)$
β_2	$U \sim (-1.5, 6)$
β_3	$U \sim (0, 2.5)$
β_4	$U \sim (-23, -14)$
β_5	$U \sim (-8, -4)$
β_6	$U \sim (-2, 0)$

1. Determine the $\boldsymbol{\beta_i}$ using the initial design (from step second) by maximizing the determinant of the information matrix. The information matrix of the proportional odds model is $\boldsymbol{X'WX}$ and the elements of matrix **W**:

$$\boldsymbol{W_i} = \begin{bmatrix} u_{i1} & -b_{i2} & 0 & \cdots & 0 \\ -b_{i2} & u_{i2} & -b_{i3} & \cdots & 0 \\ 0 & -b_{i3} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -b_{i,J-1} \\ 0 & 0 & 0 & -b_{i,J-1} & u_{i,J-1} \end{bmatrix}$$

where: $u_{it} = g_{it}^2 \left(\pi_{it}^{-1} + \pi_{it+1}^{-1} \right), \quad i = 1, \dots, 150; \quad t = 1,2$

$b_{it} = g_{it-1} g_{it} \pi_{it}^{-1}, \qquad i = 1, \dots, 150; \quad t = 1,2$

$\pi_{it} = \frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}, \qquad i = 1,2, \dots, 150; \quad j = 1,2,3$

$g_{ij} = \left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \right) \left(1 - \left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \right) \right), \qquad j = 1,2,3$

Procedure of algorithm

1. Define the candidate set of possible factor combinations from the pure component and the others point follow the Cox direction procedure for a $\delta_i = 0.025$ change in component x_i as follows:

$$\tilde{x}_i = x_i + \delta_i$$
$$\tilde{x}_j = x_j - \frac{\delta_i x_j}{1 - x_i} \,, \qquad j \neq i, i = 1,2,\ldots,p$$

2. Determine the number of the experimental runs. In this case, n = 12.
3. Select a random initial set of twelve points from the candidate set and calculate the determinant of the information matrix.
4. Generate the point-exchange algorithm by select one point form the candidate set randomly and change it to the other point from the initial design. Then, calculate the determinant of the new information matrix from the step third.
5. Consider all possible replacements, save the change that delivers the higher determinant of the information matrix. Go back to step fourth with a different point from the candidate set.
6. Return step third to fifth until 10000 repetitions.
7. The D-optimal design is a set of points that perform the maximum determinant of the information matrix.

3. Results and Discussion

Before the optimal design points obtained, the initial parameter was predicted using the Bayes method.The sets of parameters are sampled uniformly and randomly from the range of parameter estimates. The Bayes approach assumes $\boldsymbol{\beta} \sim U(a,b)$. The range of parameter estimates is shown in table 3. After simulating until 10000 of repetition, the parameter obtained with the maximum determinant of the information matrix is equal to1.89E-08. Furthermore, these parameters are used to find the optimal design points with the D-optimal criterion.

Table 3 The range of parameter estimation

Parameter	Distribution	Parameter Value	Determinant
α_1	0.00000	0.0000000	
α_2	2.00000	2.0000000	
β_1	$U \sim (3\,,5)$	4.7424160	
β_2	$U \sim (-1.5\,,6)$	2.6410900	
β_3	$U \sim (0\,,2.5)$	0.6767541	1.89E-08
β_4	$U \sim (-23\,,-14)$	-15.836790	
β_5	$U \sim (-8\,,-4)$	-4.6890930	
β_6	$U \sim (-2,0)$	-0.5413249	

Based on the parameters above, we can determine twelve design points by constructed the point-exchange algorithms. Its determinant of the information matrix is 1.588132E-14.These design points are the result of the D-optimal design of this experiments.

Table 4 The D-optimal design of ingredients

No	Component Mixed		
	X1	**X2**	**X3**
1	0.100	0.100	0.800
2	0.203	0.122	0.675
3	0.200	0.200	0.600
4	0.300	0.100	0.600

5	0.300	0.100	0.600
6	0.100	0.300	0.600
7	0.103	0.172	0.725
8	0.200	0.100	0.700
9	0.103	0.172	0.725
10	0.300	0.100	0.600
11	0.200	0.200	0.600
12	0.100	0.100	0.800

The proportion of ingredients can be described as design points as follows:
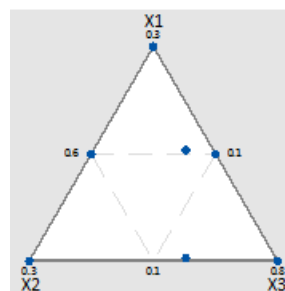

Figure 2 The The experimental region of the D-optimal design

Based on Figure 2 there are seven design points arranged. Three design points from the edges and others point in the middle of the design area.

4. Conclusion
From this experiment, there are twelve optimal design points of ingredients with seven different points. The design points that will be tested to produce cookies is all twelve optimal design. Further research can determine the optimal design points using the data that obtained from this experiment. The optimal design that generated from this data expects will be more optimal and will produce the better taste of cookies than this experiment.

5. Acknowledgment

References
[1]   Agresti A 2010 *Analysis of Ordinal Categorical Data* (New Jersey: Wiley Hoboken)
[2]   Anderson J 1984 Regression and ordered categorical variables (*Journal of the Royal Statistical Society)*46(1): 1-30
[3]   Atkinson Donev 2006 *Optimum Experimental Design with SAS* (Oxford (UK): Oxford University Pr)
[4]   Chaloner K and Verdinelli I 1995 Bayesian Experimental Design (*Review Statistical Science)* 10:273−304
[5]   Chipman H and M Hamada 1996 Bayesian analysis of ordered categorical data from industrial experiments (*Technometrics)* 38
[6]   Cornell JA1990 *Experiment With Mixtures Designs, Models And The Analysis of Mixture Data* (New York: John Wiley & Sons)
[7]   Fedorov V 1972 *Theory of Optimal Experiments* (New York: Academic Press)

[8]   Goos P and Donev A N2006The D-Optimal Design of Blocked Experiments with Mixture Components (*Journal of Quality Technology)* 38 (4): 319-332

[9]   Goos P and Jones B 2011 *Design of Experiments: A Case Study Approach* (UK: Wiley)

[10]   Gotwalt C M, B A Jones, D M Steinberg 2009 Fast computation of designs robust to parameter uncertainty for nonlinear settings (*Technometrics)* 51(1): 88-95

[11]   Mancenido Michelle V2016 Categorical Responses In Mixture Experiments [Dissertation] Arizona State University

[12]   Mancenido M, R Pan, D Montgomery 2016 Analysis of subjective ordinal responses in mixture experiments (*Journal of Quality Technology)*

[13]   Mardia K, J Kent, J Bibby 1979 *Multivariate analysis* (New York: Academic Press)

[14]   McCullagh P1980 Regression models for ordinal data (*Journal of the Royal Statistical Society)* 42(2): 109-142

[15]   Meilgaard, Moten, G V Civille, B T Carr 1999 *Sensory Evaluation Technique* (New York: CRC Press)

[16]   Meyer R K, C J Nachtsheim 1995 The Coordinate-xchange Algorithm for Constructing Exact Optimal Experimental Designs (*Technometrics)* 37(1): 60-69

[17]   Mitchell T J 1974 An Algorithm for the Construction of D-Optimal Experimental Designs (*Technometrics)* 16(2): 203-210

[18]   Montgomery, John Wiley, Sons 2001 *Design and Analysis of Experimentsfifth edition* (New York: John Wiley & Sons)

[19]   Piepel G F, S K Cooley, B Jones 2005 Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm (*Quality Engineering)* 17(4): 579-594

[20]   Ruseckaite A, P Goos, D Fok 2014 Bayesian d-optimal choice designs for mixtures (*Technical Leibniz Information Centre for Economics)*

[21]   Scheffe H 1958 Experiments with mixtures (*Journal of the Royal Statistical Society)* 20(2): 344-360

[22]   Sinha B K, Mdanal, N K Pal M, Das P 2014 *Optimal Mixture Experiments* (New Delhi: Springer)

[23]   Triefenbach F 2008 Design of experiments: The D-Optimal approach and Its Implementation As a Computer Algorithm [tesis] (Sweden (SE): Umea University dan Germany (DE): South Westphalia University of Applied Sciences)

[24]   Yang J, L Tong, A Mandal 2015 D-optimal designs with ordered categorical data (*StatisticaSinica)*