

An Approach of Factor Screening in Multifactor Mixture Experiments using Rough Set Theory

Chetan¹, Nisha²

Research Scholar, Department of Statistics, M.D. University, Rohtak, Haryana, India

Abstract: A large number of experiments are being conducted in industrial and other branches of science with an objective to compare and test quantity of inputs applied as mixture of two or more components. In particle, there are many situations where mixture experiments involve large number of factors, such types of mixture experiments are known as multifactor mixture experiments. Due to their practical applicability it is important to make such experiment user friendly. Reducing the number of factors without loss of information is an important step in this direction. Therefore, in present study rough set theory (RST), which is an extension of the set theory and has the implicit feature of compressing the dataset, is used to reduce the factors in the multifactor mixture experiments.

Keywords: Mixture Experiments, Rough Set Theory, Factor Reduction, Indiscernibility Relation, etc.

1. Introduction

Experimental design is a powerful tool for quality improvement. In some situation, however, the design variables are subject to multiple linear constraints. These situations arise quite frequently in oil industry, the alloy industry and the food industry [11, 3]. Empirical models that relate the characteristics of the products manufactured in different industries with the proportions of their ingredient are called mixture model and related designs of experiments are called mixture designs or simplex designs. Studies of this branch of science called experiments with mixtures or mixture experiments [8] and appear to have started in late fifties [6].

Mixture experiments introduced by Scheffé [6, 7] allow investigating only one factor at a time. There do occur experimental situations where proportions of components of two or more independent factors are to be tested. Designs and models for such multifactor mixture experiments were suggested by Lambrakis [5] and Nigam [1]. If a multifactor mixture experiments consist of two factors X and Z where the first factor (X) has p –components and second factor (Z) has q –components, then this two factor mixture will have the following constraints.

$$0 \leq x_i \leq 1 \quad \sum_{i=1}^p x_i = 1$$

and

$$0 \leq z_j \leq 1 \quad \sum_{j=1}^q z_j = 1$$

In general, for n –factors mixture experiment, if x_{ij} represent the j^{th} component of i^{th} factor is having p_i components, then the restrictions are

$$0 \leq x_{ij} \leq 1 \quad j = 1, 2, \dots, p_i; \\ i = 1, 2, \dots, n \\ \sum_{i=1}^n \sum_{j=1}^{p_i} x_{ij} = 1$$

Such experimental situations are known as multifactor mixture experiments.

In practice, when n is very large, it is very difficult to deal with such type of experiments. So it is important to develop some techniques to reduce the dimensionality of mixture structure with loss of generality and technique of attribute reduction in rough set theory may play an important role in such type situations. Pawlak [12] introduce the concept of RST to deal with uncertain, incomplete or vague information. RST is easy to use since it does not require additional information such as probability distribution, a prior probability etc [2]. RST is an extension of set theory and has the implicit feature of compressing the dataset. Such compression is due to definition of equivalence classes based on indiscernibility relations and to the elimination of redundant or meaningless attributes. In short, attribute reduction in RST implies in the calculation of the reducts. The classic approach is to obtain the discernibility matrix, to determine its corresponding discernibility function and to simplify it, in order to get the set of reducts. In this approach, the calculation of reducts requires the discernibility matrix. For given information system S with n elements has a symmetric discernibility matrix with dimension $n \times n$. The entries of the matrix are denoted by c_{ij} for $i \neq j$. Each entry contains the subset of attributes that distinguishes element x_i and x_j , being the diagonal entries null, according to the definition

$$c_{ij} = \{a \in A | a(x_i) \neq a(x_j)\} \text{ for } i = 1, 2, \dots, n; i \neq j$$

The corresponding discernibility function f_A is a Boolean function of m attributes (a_1, a_2, \dots, a_m) given by

$$f_A(a_1^*, a_2^*, \dots, a_m^*) = \bigwedge \{ \bigvee c_{ij} | 1 \leq j < i \leq n, c_{ij} \neq \emptyset \}$$

2. Methodology

Consider the mixture model proposed by Alam [9] for multifactor mixture experiment with n –factors and i^{th} factor is having p_i components $i = 1, 2, \dots, n$. Let x_{ij} be the j^{th} component of the i^{th} factor, $j = 1, 2, \dots, p_i$. Then the second order mixture model is given by

$$Y = \sum_{i=1}^n \sum_{j=1}^{p_i} \beta'_{ij} x_{ij} + \sum_{i=1}^n \sum_{j < j'}^{p_i} \gamma'_{ijj'} x_{ij} x_{ij'} + \sum_{i=1}^n \sum_{j, j'}^{p_i} \delta'_{iij'} x_{ij} x_{i'j'} + \varepsilon$$

Layout plan of design for multifactor mixture experiments obtained by above model using central composite design as initial design with one center point for $n = 3$ and $p_1 = p_2 = p_3 = 2$ is given in Table 1.

Table 1

| S. No. | Factor 1 | | Factor 2 | | Factor 3 | |
|--------|----------|----------|----------|----------|----------|----------|
| | x_{11} | x_{12} | x_{21} | x_{22} | x_{31} | x_{32} |
| 1 | 0.272 | 0.728 | 0.427 | 0.573 | 0.203 | 0.797 |
| 2 | 0.272 | 0.728 | 0.427 | 0.573 | 0.797 | 0.203 |
| 3 | 0.272 | 0.728 | 0.854 | 0.146 | 0.203 | 0.797 |
| 4 | 0.272 | 0.728 | 0.854 | 0.146 | 0.797 | 0.203 |
| 5 | 0.815 | 0.185 | 0.427 | 0.573 | 0.203 | 0.797 |
| 6 | 0.815 | 0.185 | 0.427 | 0.573 | 0.797 | 0.203 |
| 7 | 0.815 | 0.185 | 0.854 | 0.146 | 0.203 | 0.797 |
| 8 | 0.815 | 0.185 | 0.854 | 0.146 | 0.797 | 0.203 |
| 9 | 0.086 | 0.914 | 0.641 | 0.359 | 0.500 | 0.500 |
| 10 | 0.543 | 0.457 | 0.282 | 0.718 | 0.500 | 0.500 |
| 11 | 0.543 | 0.457 | 0.641 | 0.359 | 0.500 | 0.500 |

Now, if the experimenter wants to study the situation in which first component of first and last factor should be greater than 0.5 and first component of second factor should be greater than 0.5 and second component should be less than or equal to 0.1. Now assign the attribute Yes (Y) and No (N) to the different components of different factor according to the experimental situation under consideration. This will help us to construct the intermediate dataset with three conditional attributes.

Table 2

| S. No. | Factor 1 | | | Factor 2 | | | Factor 3 | | |
|--------|----------|----------|------------------------|----------|----------|------------------------|----------|----------|------------------------|
| | x_{11} | x_{12} | $x_{11} \wedge x_{12}$ | x_{21} | x_{22} | $x_{21} \wedge x_{22}$ | x_{31} | x_{32} | $x_{31} \wedge x_{32}$ |
| 1 | N | N | N | N | N | N | N | N | N |
| 2 | N | N | N | N | N | N | Y | Y | Y |
| 3 | N | N | N | Y | N | N | N | N | N |
| 4 | N | N | N | Y | N | N | Y | Y | Y |
| 5 | Y | Y | Y | N | N | N | N | N | N |
| 6 | Y | Y | Y | N | N | N | Y | Y | Y |
| 7 | Y | Y | Y | Y | N | N | N | N | N |
| 8 | Y | Y | Y | Y | N | N | Y | Y | Y |
| 9 | N | N | N | Y | N | N | N | Y | N |
| 10 | Y | Y | Y | N | N | N | N | Y | N |
| 11 | Y | Y | Y | Y | N | N | N | Y | N |

The discernibility matrix for the example dataset is appears as

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|----|-------|-------|-------|-------|----|----|----|----|----|----|----|
| 1 | -- | | | | | | | | | | |
| 2 | F3 | -- | | | | | | | | | |
| 3 | ∅ | F3 | -- | | | | | | | | |
| 4 | F3 | ∅ | F4 | -- | | | | | | | |
| 5 | F1 | F1,F3 | F1 | F1,F3 | -- | | | | | | |
| 6 | F1,F3 | F1 | F1,F3 | F1 | F3 | -- | | | | | |
| 7 | F1 | F1,F3 | F1 | F1,F3 | ∅ | F3 | -- | | | | |
| 8 | F1,F3 | F1 | F1,F3 | F1 | F3 | ∅ | F3 | -- | | | |
| 9 | F1 | F1,F3 | F1 | F1,F3 | ∅ | F3 | ∅ | ∅ | -- | | |
| 10 | F1 | F1,F3 | F1 | F1,F3 | ∅ | F3 | ∅ | ∅ | ∅ | -- | |
| 11 | F1 | F1,F3 | F1 | F1,F3 | ∅ | F3 | ∅ | ∅ | ∅ | ∅ | -- |

The Boolean simplification of f_A yields the set of reducts of A is given by
 $f(x) = [(F3) \wedge (F3) \wedge (F1) \wedge (F1 \vee F3) \wedge (F1) \wedge (F1 \vee F3)..]$

$$\dots \wedge [(F3) \wedge (F1 \vee F3)] \dots \wedge [(F1 \vee F3)] \wedge [(F3)]$$

This logical function is expressed by the conjunction of many terms, with each one corresponding to column of the related discernibility matrix. The Boolean simplifications of this function yields then a single reduct

$$f(x) = F1 \wedge F3$$

Therefore, this reduct is composed of factor F1 and F3 and the original dataset with three factors which can be represented in the shorter form without the loss of generality by a new dataset with two factors (Factor 1 and Factor 3)

Table 3

| S. No. | Factor 1 | | Factor 2 | |
|--------|----------|----------|----------|----------|
| | x_{11} | x_{12} | x_{31} | x_{32} |
| 1 | 0.272 | 0.728 | 0.203 | 0.797 |
| 2 | 0.272 | 0.728 | 0.797 | 0.203 |
| 3 | 0.272 | 0.728 | 0.203 | 0.797 |
| 4 | 0.272 | 0.728 | 0.797 | 0.203 |
| 5 | 0.815 | 0.185 | 0.203 | 0.797 |
| 6 | 0.815 | 0.185 | 0.797 | 0.203 |
| 7 | 0.815 | 0.185 | 0.203 | 0.797 |
| 8 | 0.815 | 0.185 | 0.797 | 0.203 |
| 9 | 0.086 | 0.914 | 0.500 | 0.500 |
| 10 | 0.543 | 0.457 | 0.500 | 0.500 |
| 11 | 0.543 | 0.457 | 0.500 | 0.500 |

Design with three factors and two factors are then compared by obtaining the G-efficiency which seeks to minimize the maximum entry in the diagonal of the hat matrix [8]. The imparical formulation of G-efficiency is given by

$$G - efficiency = \frac{p}{n \times d}$$

where n is the number of design point in the design, p is the number of parameters in the model and $d = \max[X(X'X)^{-1}X']$ over a specified set of design points.

$$G - efficiency(\text{with three factors}) = 0.9524$$

$$G - efficiency(\text{with two factors}) = 0.7964$$

For practical purpose, Wheeler [10] suggested that any design with G-efficiency $\geq 50\%$ could be good design and showed that pursuit of higher efficiencies is not generally justified in practice. The G-efficiency for both the designs is greater than 0.5 so both the designs are good for practical purpose.

3. Discussion and Conclusion

Formulation of multifactor mixture experiments are commonly found in pharmaceutical, chemical and food industries as well as in other industrial segments. In these experiments, the factors are proportions of the components in mixture and the response is a variable that characterizes the quality of the product manufactured. In general, the experiments conduct in industries involved the large number of factors. So it is very important to reduce these factors with the loss of information. For doing such dimension reduction rough set theory is very powerful problem solving technique that assists industries for tackling quality control problems effectively and economically. This paper illustrates the factor reduction technique in multifactor mixture experiments using rough set theory. The reduced design is evaluated with G-efficiency. It is found that after reducing the factors the design is still good for the practical purpose.

References

- [1] A.K Nigam, "Multifactor Mixture Experiments". Journal of the Royal Statistical Society, B35, 51-56, 1973.
- [2] A.S.A Pessoa, S. Stephany, "An Innovative Approach for Attribute Reduction in Rough Set Theory", Intelligent Information Management, pp 223-239, 2014.
- [3] Chetan, Ratna Raj Laxmi, "Diagnosis for Leverage and Influence in Mixture Experiments for Engine Management System", International Journal of Science and Research, 3(11), pp 131-134, 2014.
- [4] Chetan, Ratna Raj Laxmi, Mithlesh, "Construction of G-Optimal Multi-Factor Mixture Experiments using Fractional Factorial Designs", International Journal of Mathematics and Computer Applications Research, 6(1), pp 1-6, 2016.
- [5] D.P. Lambrakis, "Experiments with Mixtures: A Generalization of the Simplex-Lattice Design", Journal of Royal Statistical Society, B30 (1), pp 123-136, 1968.
- [6] H. Scheffé, "Experiments with Mixtures. Journal of the Royal Statistical Society, B20(2), 344-360, 1958.
- [7] H. Scheffé, "Simplex-Centroid Design for Experiments with Mixtures", Journal of the Royal Statistical Society, B25(2), 235-263, 1963.
- [8] J.A. Cornell, "Experiments with Mixtures: A Review", Technometrics, 15(3), 437-455, 1973.
- [9] N. M. Alam, "Some contribution to design and analysis of mixture experiments", Ph.D. Thesis, IASRI (IARI), New Delhi, 2010, Unpublished.
- [10] R. E. Wheeler, "Efficient experimental design. Presented at the Annual Meeting of the American Statistical Association, Montreal, Canada, 1972.
- [11] Ratna Raj Laxmi, Chetan, "Application of Experiments with Mixture in Food Product Manufacturing", International Journal of Statistika and Matematika, 10(3), pp 80-84, 2014.
- [12] Z. Palak, "Rough Set", International Journal of Information and Computer Science, 11(5), pp 341-356, 1982.