Noname manuscript No.
(will be inserted by the editor)

Algorithm to generate multi-factorial experiments to teach experimental design

Do you have a subtitle? If so, write it here

Cristy · Nagamani · Raji · S. K. Gadi

Received: date / Accepted: date

Abstract One of the challenges in teaching the subject, Design of Experiments, is to come up with a proper numerical example. In this article, authors present a methodology to generate a numerical example for multifactorial experiments. Also, it presents a simple algorithm, which can be implemented in any programming language to generate unique examples.

Keywords Experimental design; educational tool; generating examples

1 Introduction

Experimental design is applied in almost all the fields involving experimentation [1–4]. It is part of various undergraduate and graduate curriculum, ranging from the engineering to the biological sciences. In general the objective of experimental design is to minimize cost and time of the experiments and maximize the yield. Improper design of experiment may lead to inaccurate or false conclusions, as well as a loss of money, material and time [5].

Learning statistics or mathematics in general is effective by solving a number of numerical examples [6]. It helps the students to develop insight in the topics [7]. Teachers may involve students in finding experiments to teach the topic [8–10]. However, it is teacher's task to generate examples for the classroom and for the practise [11].

Solving optimization problems, finding mathematical model etc. in experimental design involves performing various experiments with a different combinations of the factors. Conducting experiments on a real system for the classroom purpose is not always feasible due to any the following limitations.

1. The cost of conducting experiments on a real system is not always negligible.
2. A considerable amount of time may take for each experiment.
3. The combination of factors associated for optimum response is constant for a physical system. Therefore, teachers may not provide a fresh problem.

Hence, a computer program generating responses for the given input factors is a good alternative to mimic the physical systems. In this article we present a methodology to generate numerical examples which simulate experiments. The objective is to generate unique process for the limits selected by the user, which ouputs experimental data for the given combinations of the factors. Teachers may adopt this methodology in generating numerical examples, which highlight all the characteristics they want to present to the classroom, give as practice exercise and conduct exams.

The proposed algorithm is described in Section 4. Readers interested only in the implementation of algorithm may skip the mathematical construction presented in Section 2 and 3.

A numerical example for an experimental design is a mathematical model representing a physical process.

Cristy
first address

Nagamani
first address

Raji
second address

S. K. Gadi
FIME, Universidad Autónoma de Coahuila, Torreón, México
Tel.: (+52 1) 871-757-0153
Fax: (+52 1) 871 757 0153
E-mail: research@skgadi.com

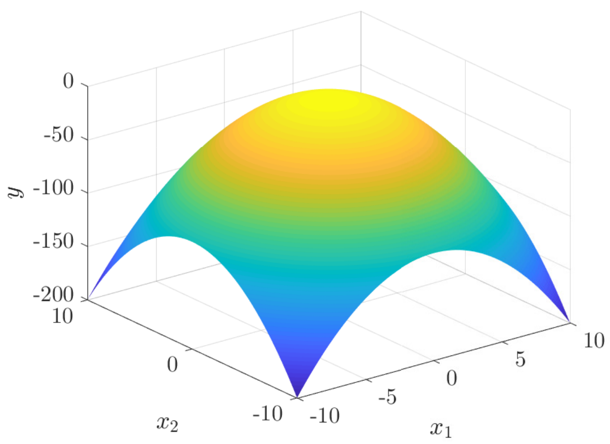

Fig. 1 A second order polynomial convex function

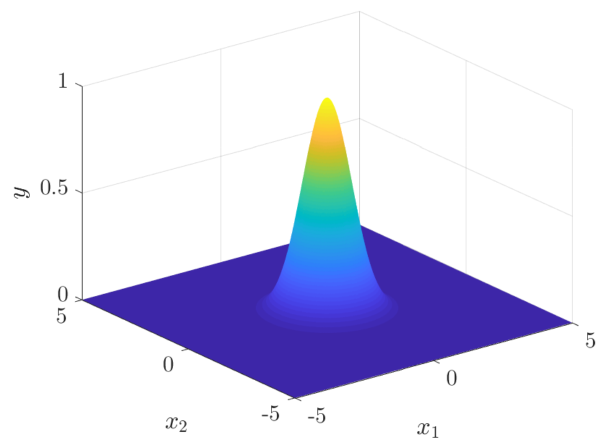

Fig. 2 Two variable Gaussian distribution function

This model is a set of static functions (i.e. it does not have derivative or integral terms) which maps the factors to the responses. A real life system may present more than one peaks. However, most of the experimental design methods find the local maximum based on the initial base value. Hence, the proposed algorithm is designed to present only one peak. A multi-response system can be represented as

$$y_j = f_j(x_1, x_2, x_3, \ldots, x_n) + \xi_j \tag{1}$$

where y_j, $j \in \{1, 2, 3, \ldots, m\}$ are the responses, x_i, $i \in \{1, 2, 3, \ldots, n\}$ are the factors, f_j, $j \in \{1, 2, 3, \ldots, m\}$ are the nonlinear functions mapping the n factors to the m responses and ξ_i, $i \in \{1, 2, 3, \ldots, m\}$ are the noise.

All the factors, x_i, are constrained by upper and lower limits. The numerical examples should produce an unique optimal responses, y_j^M, for a set of factors within its limits. Construction of a one such mathematical function is presented in the next section.

The proposed algorithm presents the case of single response, which can be adopted to multi-response.

2 Construction of a mathematical function to suit our requirements

2.1 Quadratic concave function

A second order polynomial function, such as

$$F(x_1, x_2, x_3, \ldots, x_n) = -\sum_{i=1}^{n} x_i^2 \tag{2}$$

is a concave function, which serve the purpose of providing a unique optimal point. Figure 2.1 depicts (2) for the two variables case. However, it doesn't meet the requirements of a good example because to the following limitations.

1. Response surface methodology uses a second order fit algorithm. Hence, the process of reaching optimal solution becomes trivial.
2. A quadratic function is having a property that its slope increases as it moves far from the optimal point. This property trivializes the process of selecting a new base value.

2.2 Multivariable Gaussian function

The multivariable Gussian function

$$F(x_1, x_2, x_3, \ldots, x_n) = \prod_{i=1}^{n} \mathrm{e}^{-x_i^2} \tag{3}$$

is a concave function, hence, it has a unique maximum value. The slope of this function is not linearly related with the distance from its optimal point. The concave functions have property that the response of all the points between any two arbitrary points always grater than the responses at these arbitrary points [12]. A non-concave function gives additional challenge in solving the optimization problem.

2.3 Sigmoid convex function

Keeping above limitations in mind, a sigmoid based convex function is proposed. One variable sigmoid function is

$$S(x) = \frac{1}{1 - \mathrm{e}^{-x}} \tag{4}$$

and its derivative is

$$F(x) = S'(x) = \frac{\mathrm{e}^{x}}{(\mathrm{e}^{x} + 1)^2} = S(x)(1 - S(x)). \tag{5}$$

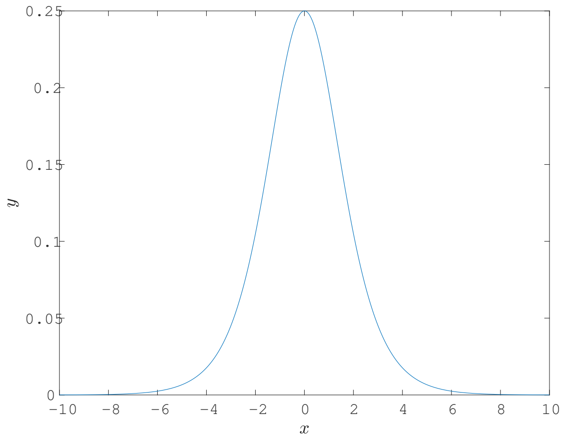

Fig. 3 One variable convex function

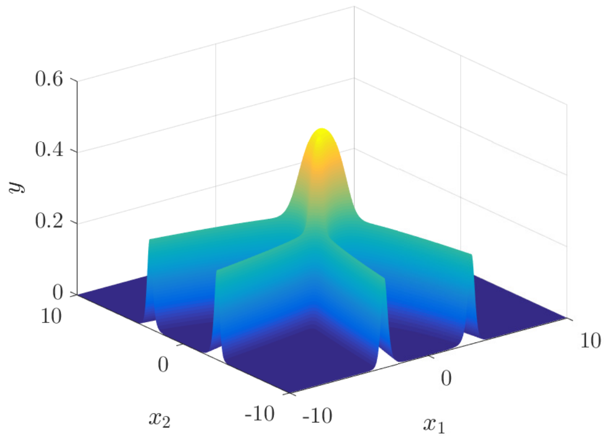

Fig. 5 Two variable function future

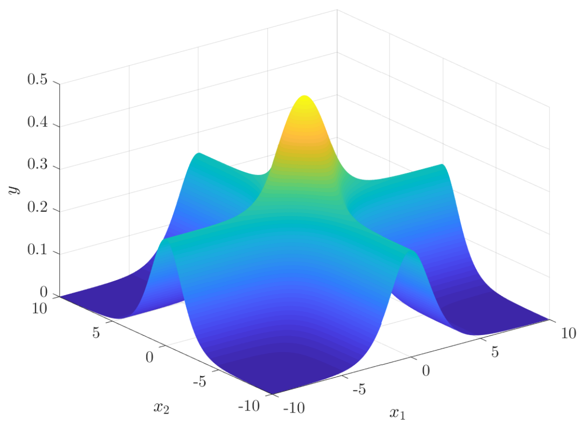

Fig. 4 Two variable convex function

The function F is plotted in Figure 2.3. The function F can be extended to a n variables case as

$$F(x_1, x_2, x_3, ..., x_n) = \sum_{i=1}^{n} \frac{e_i^x}{(e^{x_i}+1)^2}. \tag{6}$$

Figure 2.3 shows sigmoid based function for two variables. A function can be said concave, if Hessian matrix associated with it is negative definite [13]. Hessian matrix for (6) is

$$H_{i,j} = \frac{\partial^2 F}{\partial x_i^2} = \begin{cases} \frac{e^{x_i}(e^{2x_i}-4e^{x_i}+1)}{(e^{x_i}+1)^2}, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

It can be observed that the Hessian matrix, $H_{i,j}$, is not a negative definite because this diagonal matrix can be positive for $x_i < \ln(2-\sqrt{3})$ and $x_i < \ln(2+\sqrt{3})$

The Hessian matrix, $H_{i,j}$ is having a not a negative definite because it is a diagonal matrix and the diagnal values are positive. Hence, F is a convex function, i.e. has a unique maximum value with no other peaks. In the next section a method is presented to adapt F to generate random experiments within given limits.

Not using normal distribution because we can tune this further to get a non-symmetric equation.

3 Adapting convex function

It can be observed that the convex function, F, proposed in the previous section has a maximum value $F_{max} = 0.25n$ at $x_i = 0, \forall i \in \{1, 2, 3, \ldots, n\}$, where n is number of factors. Also, it can be noted that the response F reaches close to zero when x_i

x_i^L and x_i^U are the lower and upper limit of a factor. Let X_i^M be the values where F obtains maximum value.

$$F(x_1, x_2, x_3, ..., x_n) = \sum_{i=1}^{i=n} \frac{e^{(x_i - x_i^M)}}{[e^{(x_i - x_i^M)} + 1]^2}. \tag{8}$$

A maximum value of F_M is obtained by taking

$$F(x_1, x_2, x_3, ..., x_n) = \frac{4F_M}{n} \sum_{i=1}^{i=n} \frac{e^{(x_i - x_i^M)}}{[e^{(x_i - x_i^M)} + 1]^2}. \tag{9}$$

The maximum value, F_M, can be obtained by

$$F_M = F_L + \xi(F_U - F_L) \tag{10}$$

where F_L is lower limit, F_M is upper limit ξ is a constant random value, which helps to generate a new F_M ever time the above function is invoked.

The range of interest is between x_i^L and x_i^U. A maximum value should be between these limits. Also, it is recommended not to move maximum values to the extremes. Hence we propose

$$x_i^M = x_i^L + \frac{\xi(1-\alpha)(x_i^U - x_i^L)}{2} \tag{11}$$

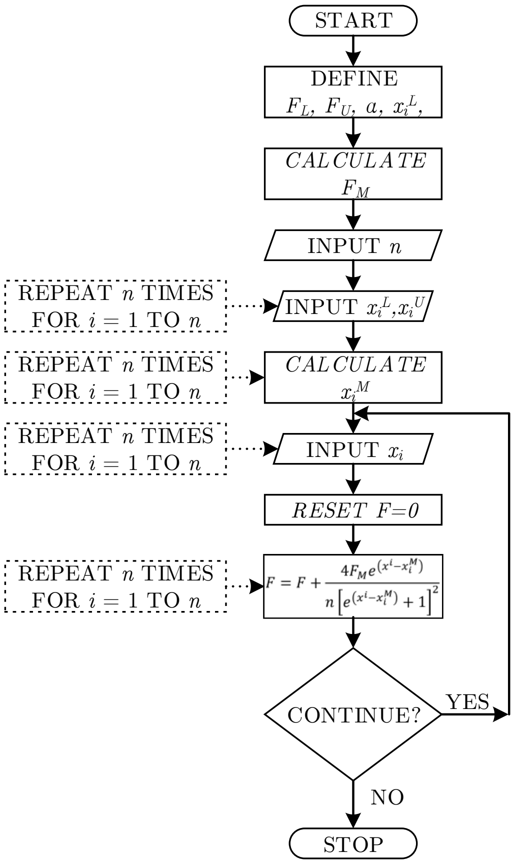

Fig. 6 Flowchart

where $0 < \alpha < 1$ is a constant and $0 < \xi \leq 1$ is a constant random number. α determines the region where maximum value may fall. The random value, ξ, allows to select a new point each time this function is invoked.

4 Algorithm

Figure 4 shows the proposed algorithm.

5 Application

6 Conclusion

The mathematical construction is given in detail, which helps others to adapt to meet any additional requirements.

References

1. R. A. Fisher, *The design of experiments.* Oliver And Boyd; Edinburgh; London, 1937.
2. G. Quinn and M. Keough, *Experimental Design and Data Analysis for Biologists.* Cambridge University Press, 2002.
3. D. Montgomery, *Design and Analysis of Experiments.* Student solutions manual, John Wiley & Sons, 2008.
4. J. Antony, *Design of Experiments for Engineers and Scientists.* Elsevier insights, Elsevier Science, 2014.
5. M. F. Festing, "Principles: The need for better experimental design," *Trends in Pharmacological Sciences*, vol. 24, no. 7, pp. 341 – 345, 2003.
6. X. Zhu and H. A. Simon, "Learning mathematics from examples and by doing," *Cognition and instruction*, vol. 4, no. 3, pp. 137–166, 1987.
7. A. Renkl, "Learning from worked-out examples: A study on individual differences," *Cognitive science*, vol. 21, no. 1, pp. 1–29, 1997.
8. W. G. Hunter, "Some ideas about teaching design of experiments, with 25 examples of experiments conducted by students," *The American Statistician*, vol. 31, no. 1, pp. 12–17, 1977.
9. M. N. Fried, "Mathematics as a constructive activity: Learners generating examples," *ZDM*, vol. 38, no. 2, pp. 209–211, 2006.
10. S. M. Hiebert, "Teaching simple experimental design to undergraduates: do your students understand the basics?," *Advances in Physiology Education*, vol. 31, no. 1, pp. 82–92, 2007.
11. D. L. Ball, M. H. Thames, and G. Phelps, "Content knowledge for teaching," *Journal of Teacher Education*, vol. 59, no. 5, pp. 389–407, 2008.
12. A. Antoniou and W.-S. Lu, *Practical optimization: algorithms and engineering applications.* Springer Science & Business Media, 2007.
13. B. Bernstein and R. Toupin, "Some properties of the hessian matrix of a strietly convex function.," *Journal für Mathematik. Bd*, vol. 210, no. 1/2, p. 9, 1962.