A computer-based educational tool for generating experimental design examples

# Abstract

In this article, challenges in designing example problems of multifactorial experiments for an academic purpose are identified, a mathematical function is proposed to mimic a physical system, and an algorithm for generating multifactorial experiments is proposed. The proposed mathematical has the following properties: 1) It can be adapted for any number of finite factors; 2) It contains a unique peak; 3) It is not a convex function; and 4) (Yo no se como explicar eso. Por favor mejora eso basando el punto 4 que esta abajo de las equaciones de C\_1, C\_2 y C\_3) There is no linear relation between the partial derivatives of the function with respect to its factors. The proposed algorithm has the following features: 1) It employs a proposed mathematical function; 2) 2) It allows the students to simulate the behavior of multifactorial experiments; 3) It generates functions with random for uniqueness. An application is developed based on this algorithm with the help of JavaScript, HTML and CSS. This application is tested in teaching Response Surface Methodology.

# Keywords

Experimental design, educational tool, generating examples, problem-based learning

# Introduction

Almost all the fields involving experimentation use experimental design [1]–[4]. It is part of various undergraduate and graduate curriculum, ranging from engineering to biological sciences. The objective of experimental design is to minimize cost and time of experiments and maximize its yield. As an example, it can be used to find the values of factors such as pH, oxygen concentration, sugar concentration, with which the enzyme production is maximized. Different techniques can be employed to find the minimum number of experiments [5] . On the other hand, an improper design of experiment may lead to inaccurate or false conclusions, as well as a loss of money, material and time [6].

Solving many numerical examples helps students to learn statistics or mathematics in general [7] , and to develop insight into the topics [8] . Moreover, students show good learning experience using visual examples and perform better with examples of experiments, which they can relate [9] . On the other hand, solving optimization problems and finding the most accurate mathematical model for a process in experimental design involves performing various experiments with different combinations of factors. Conducting experiments on a real system is not always feasible due to any of the following limitations.

1. Experiments on a real system can be costly.
2. A considerable amount of time may take each experiment.
3. The values for the factors causing optimum-response to a physical system are constant. Hence teachers may not provide a great variety of example problems.

Hence, a computer program generating responses for some specific inputs is an excellent alternative to mimic the physical systems. In this article, a methodology is presented to generate numerical examples, which simulate experiments. The objective is to generate a unique process that takes into account factor limits selected by the user, which outputs experimental data for a given combinations of parameters. Teachers may adopt this methodology in order to generate numerical examples, which highlight all the characteristics they want to present to students. Teachers can use the proposed methodology for evaluating the knowledge acquired by students during the course. Moreover, this technique allows teachers to implement problem-based learning. In this pedagogy, a student learns the topic while solving a problem given by the teacher.

Note that a numerical example for an experimental design is a mathematical model representing a physical process. 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. The response of a physical system may present more than one local maximum. 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.

Let a multi-response system given by

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

where ,  are responses, ,  are factors, ,  are nonlinear functions mapping the  factors to the  responses and ,  are zero mean random noise.

All the factors, , are constrained by upper and lower limits. The numerical examples should produce unique optimal responses, , for a set of factors within its limits.

The proposed algorithm presents the case of single response. A set of  singe response systems can mimic a multi-response system.

The following section presents the mathematical functions which can simulate a real-life process with the following properties.

1. There exists a unique maximum value.

2. The functions are static. Time can be a factor but does not use the derivative and integral terms.

# Proposed mathematical function

A static mathematical function with a single peak is achieved by any of the following functions.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | (2) | |
|  |  | (3) | |
|  | . | (4) |

The functions  and are a quadratic polynomial, multivariable Gaussian function respectively. The function is obtained by replacing the multiplication operator with a summation operator. Figures 1, 2 and 3 respectively represent the functions , , and  for a two variable case. However, these functions have the following limitations:

1. The function  has 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. The functions  and are convex function. The concave functions have a property that the response of all the points between any two arbitrary points always greater than the responses at these arbitrary points [14]. This is not recommended because this property also makes the selection of next base value easy.

3. Response surface methodology uses a second order fit algorithm. Hence, reaching optimal solution for a system based on  requires very less effort, which is not recommended as a practice problem.

4. The , , and  are linear. Multi factorial example problems based on these functions are easy to solve because optimizing one factor after another factor will work in these cases.

Adding a nonlinear term to a  function gives the following proposed novel mathematical function.

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

Figure 4 depicts  for a two variable case with . The gradient of  is

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

where . Peaks, dips or saddle points are formed at

|  |  |  |
| --- | --- | --- |
|  | . | (9) |

The union of the solutions to the following three equations gives the solution to (9).

|  |  |  |
| --- | --- | --- |
|  |  | (10) |
|  |  | (11) |
|  |  | (12) |

The following procedure identifies the requirements to ensure the existence of one single peak by eliminating the occurrence of other peaks, dips or saddle points. Note that solution for (10) is , which can don't affect the solution to (9). On the other hand, the solution for (12) is

|  |  |  |
| --- | --- | --- |
|  | . | (13) |

Selecting a value , such that , all the real solutions of (13) can be suppressed.

Equation (11) can be rewriting as

|  |  |  |
| --- | --- | --- |
|  | , | (14) |

This expression implies that the solution to (9) lies at . Equation (14) is equivalent to

|  |  |  |
| --- | --- | --- |
|  | . | (15) |

Since  for all the values of  except for , the solution to (9) can be limited to only the point  provided that . By considering a positive value for , only one peak at origin guaranteed when .

Properties of the proposed mathematical function:

1. Its Gradient is not proportional to the distance from its optimum combination.
2. It has a unique maximum value at  i.e. at , provided that 
3. It is not a concave function.
4. The optimal value of an arbitrary factor is not constant throughout the factorial space.

In the next section, a method is presented to adapt the function  defined at (7) in order to generate random experiments.

# Adapting the proposed function

A scaled version of the proposed function  in (7) is

|  |  |  |
| --- | --- | --- |
|  |  | (16) |
|  | , | (17) |
|  | , | (18) |
|  | , | (19) |
|  | , | (20) |
|  | , | (21) |
|  | , | (22) |
|  | , | (23) |
|  | , | (24) |
|  |  | (25) |
|  |  | (26) |

With ,  is a random variable such that  and , where  denotes expected value. Moreover,  is a random variable with the properties  and . On the other hand  is the function range,  is a noise factor,  is difficulty factor,  are the optimal combination of factors where the function  reaches its maximum value,  and  are lower and upper limits of the  factor,  and  are padding constants for limiting the maximum value of the function  and limiting the optimal combination within the desired region respectively.

It is worth mentioning that function  shown in (16) preserves all the following mathematical properties of the function .

In the next section, an algorithm that implements the proposed procedure is presented.

# Algorithm

The objective of the algorithm is to generate experiment results by simulating a multi-factorial process. The teacher gives the constants required by the multi-factorial process, and students are allowed to perform experiments at different factors. The algorithm generates responses for the given  inputs using the equations and inequalities given from (16) to (26). The objective for the student is to minimize the number of experiments conducted to achieve the optimum value of the process.

Figure 5 shows the flowchart of the proposed algorithm. The teacher should define the values of  and , where  and  should be less than 0.5. Based on our experience, It is recommended to use .  is the difficulty factor, the bigger the value is assigned, the harder it is to reach the optimum value. It is recommended to use a value . The noise factor  introduces noise into the system.

This algorithm can be implemented in any programming language to distribute the students.

# Application

The triad of JavaScript (JS), Hypertext markup language (HTML) and Cascading Style Sheets (CSS) is used to create an application [16]. JS is used to implement the algorithm discussed in the previous section.

This application is executed with , , , , , , , , , and . The application calculated , , , , , , and . Figure 6 and 7 show, respectively, the contour plot and surface plot of the function  without the noise term i.e., considering . The RSM is applied to find the optimal values of  and , results of each iteration are also superimposed over the contours.

The master students of Biological Sciences Faculty, Universidad Autónoma de Coahuila, Torreón were instructed using this application. This tool became part of the classroom as well as the homework in the learning of Response Surface Methodology (RSM). The application does not give access to the optimum values. A distance tool is available in this application to provide the information how far is the optimum value from any set of given factors. This distance tool lets the students notice if they have arrived the optimum result.

This technique allowed the teacher to implement the problem-based learning. Also, individualize the problem, i.e., the teacher assigns a unique problem to each student. It helped the students to work in groups to discuss the technique at the same time every student has to work on themselves to solve his/her unique problem. The teacher also reported that the students became more interactive in the classroom.

# Conclusion

The Construction of a single response, unique peak multivariable mathematical function is proposed in this article. This article presents four mathematical functions (, , , and) which are suitable to mimic a real-life process for the experimental design purpose. The limitations of $C\_1$, $C\_2$, and $C\_3$ ove the proposed $C\_4$ are studied. The functions are ordered in the order of difficulty with being the easiest and  being the hardest. The function, , is adapted to generate experimental data for a selected range of factors. This article also presents an algorithm to generate experimental results. This algorithm generates a random process based on the constants provided by the teacher. An application of the algorithm is designed in HTML, CSS, and JavaScript. It is used in the classroom to teach the topic of Response Surface Methodology (RSM). However, Teachers may use this application to teach any other optimization techniques such as Taguchi methods.

The function developed in this article has only one peak, i.e., the optimal value is a maximum value. Multiplying the function, , by a negative number transforms the single-peak function to a single-dip (opposite to a peak) function. It also requires to modify (17) and (18) such that the upper and lower limits are respected.

This work can further be extended to a multiple response case by generating  number of functions  where , which requires to generate  number of values for , i.e. the values  are replaced by  where  and .

# References

[1] R. A. Fisher, *The design of experiments*. Oliver And Boyd; Edinburgh; London, 1937.

[2] G. P. Quinn and M. J. Keough, *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, 2002.

[3] D. C. Montgomery, *Design and Analysis of Experiments*. John Wiley & Sons, 2008.

[4] J. Antony, *Design of Experiments for Engineers and Scientists*. Elsevier Science, 2014.

[5] A. M. Sarotti, R. A. Spanevello, and A. G. Suarez, “An efficient microwave-assisted green transformation of cellulose into levoglucosenone. Advantages of the use of an experimental design approach,” *Green Chem.*, vol. 9, no. 10, pp. 1137–1140, 2007.

[6] M. F. W. Festing, “Principles: The need for better experimental design,” *Trends Pharmacol. Sci.*, vol. 24, no. 7, pp. 341–345, 2003.

[7] X. Zhu and H. A. Simon, “Learning mathematics from examples and by doing,” *Cogn. Instr.*, vol. 4, no. 3, pp. 137–166, 1987.

[8] A. Renkl, “Learning from worked-out examples: A study on individual differences,” *Cogn. Sci.*, vol. 21, no. 1, pp. 1–29, 1997.

[9] J. Hattie and G. C. R. Yates, *Visible Learning and the Science of How We Learn*. Taylor & Francis, 2013.

[10] W. G. Hunter, “Some Ideas about Teaching Design of Experiments, with 25 Examples of Experiments Conducted by Students,” *Am. Stat.*, vol. 31, no. 1, pp. 12–17, 1977.

[11] M. N. Fried, “Mathematics as a constructive activity: Learners generating examples,” *ZDM*, vol. 38, no. 2, pp. 209–211, 2006.

[12] S. M. Hiebert, “Teaching simple experimental design to undergraduates: do your students understand the basics?,” *Adv. Physiol. Educ.*, vol. 31, no. 1, pp. 82–92, 2007.

[13] D. L. Ball, M. H. Thames, and G. Phelps, “Content Knowledge for Teaching,” *J. Teach. Educ.*, vol. 59, no. 5, pp. 389–407, 2008.

[14] A. Antoniou and W.-S. Lu, *Practical optimization: algorithms and engineering applications*. Springer Science & Business Media, 2007.

[15] B. Bernstein and R. A. Toupin, “Some Properties of the Hessian Matrix of a Strietly Convex Function.,” *J. f{ü}r Math. Bd*, vol. 210, no. 1/2, p. 9, 1962.

[16] S. K. Gadi, “Multifactorial experiment simulator - Suresh Kumar Gadi,” 2017. [Online]. Available: https://skgadi.com/tools/multifactorial-experiment-simulator/. [Accessed: 27-Aug-2017].