A computer-based educational tool for generating experimental design examples

# Abstract

Experimental design (ED) subject deals with optimization, i.e., maximizing yield or minimizing cost by controlling various factors. Designing multifactorial experiments for an academic purpose is not always an easy task. In this article, authors identified challenges in designing the example problems. Authors provide a mathematical framework for generating various types of experiments. Based on the requirements, teachers may adopt any of the presented methods.

This article also presents an algorithm based on one of the proposed frameworks. JavaScript along with HTML and CSS is used to implement the proposed algorithm. This platform is used to teach the master students of biological sciences. Teachers used the platform in the classroom to apply problem-based learning. They observed that the number of doubts asked during the class hours has increased, hence the student's participation.

# 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 the engineering to the biological sciences. The objective of experimental design is to minimize cost and time of the experiments and maximize the yield. As an example, it can be used to find the values of the factors (such as pH, oxygen concentration, sugar concentration) for which enzyme production is maximum. Different techniques can be used 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 to learn statistics or mathematics in general [7] . It helps the students to develop insight into the topics [8] . Students show good learning experience using visual examples and perform better with the examples of experiments which they can relate [9] . Teachers may involve students in finding experiments to teach the topic [10]–[12] . However, it is teacher's task to generate examples for the classroom and the practice [13].

Solving optimization problems and finding the most accurate mathematical model for a process/system in experimental design involves performing various experiments with different combinations of the factors. Conducting experiments on a real system for the classroom purpose is not always feasible due to any of 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 with an optimum-response is constant for a physical system. Therefore, teachers may not provide a new problem.

Hence, a computer program generating responses for the given input factors 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 for the limits selected by the user, which outputs 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.

Also, 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. So, it is teacher's responsibility to design the examples such that they help to develop all the skills that are intended for the student to learn. Sometimes teachers may need to generate several problems to achieve it. Failing which a student may acquire incorrect intuition (or insight).

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. 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.

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

is a multi-response system where ,  are the responses, ,  are the factors, ,  are the nonlinear functions mapping the  factors to the  responses and ,  are the noise.

All the factors, $x\_i$, are constrained by upper and lower limits. The numerical examples should produce unique optimal responses, $y\_j^M$, for a set of factors within its limits.

The proposed algorithm presents the case of single response. If a multi-response system is required, a set of single response systems represent.

# 2. Mathematical functions

## 2.1 Quadratic concave function

# Mathematical functions

In this section different mathematical functions and its properties are presented. These properties help teachers to choose the required mathematical function.

## Quadratic concave function

A second order polynomial function, such as

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

is a concave function, which serves the purpose of providing a unique optimal point at , . Figure 1 depicts (2) for the two variables case.

Properties of quadratic concave function:

1. The quadratic concave function is a concave 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].
2. Response surface methodology uses a second order fit algorithm. Hence, the process of reaching optimal solution becomes trivial.
3. A quadratic 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.
4. The optimum value for any  factor is unaffected by the other factors.

## Multivariable Gaussian function

The multivariable Gaussian function

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

is a concave function, with a unique maximum value at , . Figure 2 shows a two variable Gaussian function.

Properties of Multivariable Gaussian function:

1. The slope of this function is not linearly related to the distance from its optimal point.
2. The multivariable Gaussian function is a concave function.
3. The optimum value for any  factor is unaffected by the other factors.

## 2.3. Modified Gaussian function

In this article, we define the modified Gaussian function as

|  |  |  |
| --- | --- | --- |
|  | . | (4) |

Figure 3 depicts (4) for the case of two variables. A symmetric matrix is negative definite when all its eigenvalues are negative. A function can be said concave if Hessian matrix associated with it is negative-definite [15]. Hessian matrix for (4) is

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

where .

The above equation shows that the Hessian matrix, , is a diagonal matrix. In a diagonal matrix, each element on the principal diagonal is an eigenvalue. So, this matrix is not a negative definite because there exist positive elements for .

In a function, a gradient is zero at the peaks, dips and saddle points. The gradient vector of (4) is

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

where .  implies  or . Hence, it is guaranteed that there exists only one peak at .

Properties of the Modified Gaussian function:

1. The slope of this function is not linearly related to the distance from its optimal point.
2. The multivariable Gaussian function is not a concave function.
3. The optimum value for any  factor is unaffected by the other factors.

## 2.4. The proposed mathematical function

Adding a nonlinear term to the modified Gaussian 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 form at

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

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

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

The solution for (10) is , which can don't affect the solution. The solution for (12) is

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

Selecting a value , all the real solutions of (13) can be suppressed. Rewriting (11) as

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

implies that the solution lies at . Hence (14) can be rewritten it as

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

Since  for all the values of  except for , the solution can be limited to only one point  provided that . Considering a positive value for , only one peak at origin for  is guaranteed.

Properties of the proposed mathematical function:

1. Gradient is not proportional to the distance from its optimum combination.
2. Has 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) to generate random experiments.

# Adapting the proposed function

A scaled version of the proposed function $C\_4$ in (7) is

|  |  |  |
| --- | --- | --- |
|  | \[\begin{align}  & f({{x}\_{1}},{{x}\_{2}},{{x}\_{3}},\ldots ,{{x}\_{n}})=F({{z}\_{1}},{{z}\_{2}},{{z}\_{3}},\ldots ,{{z}\_{n}}):= \\  & {{F}\_{1}}+\frac{{{F}\_{2}}}{n}\left[ \sum\limits\_{i=1}^{n}{\exp \left[ -{{\left[ {{z}\_{i}}+a\sin \left( {{\Sigma }\_{z}} \right) \right]}^{2}} \right]}+{{K}\_{R}}\xi \right],  \end{align}\] | (16) |
|  | \[F\_1:= F\_L+\alpha F\_I\Xi\], | (17) |
|  | \[F\_2 := (1-2\alpha)F\_I\Xi\], | (18) |
|  | \[F\_I := (F\_U-F\_L)\], | (19) |
|  | \[\Xi := 0.5 (\xi + 1)\], | (20) |
|  | \[\Sigma\_z := \sum\_{i=1}^{n} z\_i\], | (21) |
|  | \[z\_i := \frac{K\_D(x\_i-x\_i^M)}{x\_i^I}\], | (22) |
|  | \[x\_i^I := x\_i^U - x\_i^L\], | (23) |
|  | \[x\_i^M := x\_i^L + \beta x\_i^I+(1-2\beta)x\_i^I\Xi\], | (24) |
|  | \[0 < \alpha < 0.5\], | (25) |
|  | \[0 < \beta < 0.5\], | (26) |

where $i \in \{1, 2, 3, \dots, n\}$, $\xi$ is a random variable with the properties $E(\xi)=0$ and $|\xi|\le 1$, $\Xi$ is a random variable with the properties $E(\Xi)=0.5$ and $0 \le \xi \le 1$, $E(.)$ is the expected value. $[F\_L, F\_U]$ is the function range, $K\_R$ is a noise factor, $K\_D$ is difficulty factor, $x\_i^M$ are the optimal combination of factors where the function $g$ reaches its maximum value, $x\_i^L$ and $x\_i^U$ are lower and upper limit of the $i^{\text{th}}$ factor, $\alpha$ and $\beta$ are padding constants for limiting the maximum value of the function $f$ and limiting the optimal combination within the desired region respectively.

The function $F(z\_1, z\_2, z\_3, \dots, z\_n)$ of (16) preserves all the following mathematical properties of the function $C\_4$.

In the next section, an algorithm is presented to show implementation procedure.

# 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. 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 . The values of  and  should be less than 0.5. It is recommended to use .  and  are lower and upper limits of the values generated in the experiments. Hence, a user should select the values such that .  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 shows the contour plot of the function  generated with these constants. 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 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 for is presented. Later it is adapted to generate experimental data for a selected range of factors. An algorithm is proposed, which can be realized in any programming language. Based on this algorithm an application is designed in HTML, CSS and JavaScript. It is used in the classroom to teach the topic of Response Surface Methodology (RSM).

It is developed for maximum values, but can be adapted to the minimum by putting negative to the function and scaling accordingly. That is a unique dip (opposite to a peak) can be obtained by selecting a negative values for . It requires to modify (17) and (18).

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 .

A non-concave function gives an additional challenge in solving the optimization problem.

This property is not recommended because it trivializes the multi-factorial problems.

# 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].