Experimental Sequential Designs for Logistic Regression Models

Diseños experimentales secuenciales para modelos logísticos de regresión

Arturo T. De Zan[a]

Departamento de Ingeniería, Facultad de Ingeniería, Universidad de la Sabana, Chía, Colombia

Abstract

When the usual hypotheses of normality and constant variance do not hold (*e.g.* in binomial or Bernoulli processes), the problem of choosing appropriate designs creates problems to researches when pursuing a sequential exploration of process. This paper is based on De Zan (2006), where the author proposes two criteria to evaluate design strategies, that take the amount of information as the main evaluation tool. One into account the information of the fitted model, and the other explores the information that is contained on the approximation of a set of the best conditions of factors found on a fitted model. An example of how these strategies work is also given through a simulation using R software.

Key words: Factorial Design, Response Surface Design, Sequential Design of Experiments, Generalized Linear Model, Logistic Regression, Fisher Information Matrix.

Resumen

Cuando los supuestos habituales de normalidad y varianza constante no se cumplen (*e.g.* en procesos de Bernoulli o binomiales), el problema de la elección de diseños adecuados ocasiona cierta dificultad a los experimentadores, especialmente cuando lo que se persigue es una exploración secuencial del proceso. Este artículo está basado en De Zan (2006), en donde se proponen dos criterios para evaluar estrategias de diseño. Una de ellas toma en cuenta la cantidad de información contenida en el modelo ajustado, mientras que la otra explora la información contenida en las mejores condiciones de experimentación encontradas en el modelo ajustado. Se desarrolla un ejemplo simulado con el paquete R acerca de cómo funcionan estas estrategias.

Palabras clave: diseño factorial, metodología de superficie de respuesta, diseño de experimentos secuenciales, modelo lineal generalizado, regresión logística, matriz de información de Fisher.

---

[a]Assistant Professor. E-mail: arturo.de.zan@unisabana.edu.co

1. Introduction

1.1. Classical Response Surface Methodology

Since the very seminal work of Box & Wilson (1951), and for more than five decades, many researchers and practitioners of Response Surface Methodology (RSM) have widely contributed to the natural development of this matter. RSM has achieved special acceptance within industrial applications, quality improvement of processes and industrial processes in general, among others.

Four complete references that reinforce the extensive use of these applications are: Box & Draper (1987), Box & Draper (2007), Khuri & Cornell (1996) and Myers & Montgomery (2002). Some of the most recent updated aspects on this topic are also described in Myers (1999), Myers et al. (2004) and Atkinson (2006). Future lines of research are well detailed there, as well.

Although RSM problems have been developed in a very cohesive manner complete way, the scope of the majority of contributions are focused on the normal linear model[1]. The study of processes with abnormal distributions in the context of RSM is a relatively new problem, especially when some considerations related to the choosing of "good" designs arise. There are many contributions whose focus is on the *analysis* of experiments –strictly speaking– while the *design* does not seem to reach the same level of development (for instance Myers & Montgomery (2002), Myers et al. (2001a), Myers et al. (2004) and Khuri (2006) for more details).

1.2. Non-normal Models and GLMs

When the classical hypotheses of statistical modeling do not apply, especially when the usual assumptions of normality and constant variance do not hold, many problems for which there are several solutions from the point of view of experimental designs arise.

One of the first attempts to cope with this departure from the usual hypotheses is to transform the data (*e.g.* using Box-Cox family of transformations). These transformations work adequately when a specific hypothesis is tested, in the sense that a transformed variable could fit better to the classical approach of design and analysis than if it were not transformed. Nevertheless, some authors point out that there is no one transformation that can cope completely with all of the deviations from the usual scope at the same time. Also, it is said that sometimes many transformations do not make sense from a physical point of view, or may carry on some additional problems when analyzed statistically. For further details see Montgomery (2005) and Myers et al. (2001b).

The seminal work of Nelder & Wedderburn (1972) was the beginning of a new well-integrated approach for the problem of non-normal distributions, namely the Generalized Linear Models (GLMs). Several authors have later contributed to extend the applications of these models. Among the most referenced works there

---

[1]See for example Khuri & Mukhopadhyay (2006).

are: Lindsey (1997), McCullagh & Nelder (1989), Dobson (2002) and Myers et al. (2001*a*).

The result of this endeavor was successful. A new class of statistical models now can be considered and analyzed from a common and solid point of view, which is well accepted in the context of modern statistical modeling. In Myers et al. (2001*a*) the authors use the GLM approach to analyze given dataset for binomial and Poisson factorial designs. Other valuable contributions –see Lewis et al. (2001*b*) and Lewis et al. (2001*a*)– focus on some implications of analysis as well as, for example, how to obtain confidence intervals for the parameters on factorial designs when the response follows binomial and Poisson distributions. When the main goal of a study is not the *analysis*, but in the *design* of experiments, it is well known that the choice of design depends on the unknown parameters. See Cox & Reid (2000), Khuri & Mukhopadhyay (2006) and Atkinson (2006) for more details.

Compared with classical regression models (*e.g.* normal linear models), the random component of GLMs is characterized by a certain relaxation of the assumption of additive error of the former. For a n-dimensional vector of independent observations of the response, say $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, the usual model has the form $Y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$, where $\mathbf{x}$ is a vector of known explanatory variables, $\boldsymbol{\beta}$ a vector of unknown parameters and ε the error term. In this case, the density function of Y is $f_Y(y) = f_\varepsilon(y - \mathbf{x}'\boldsymbol{\beta})$ (Firth 1991). In the case of GLM, errors do not necessarily follow a normal distribution but belong to the exponential family distributions. So, the density function of the response can be written as $f_Y(y) = f(y; \mathbf{x}'\boldsymbol{\beta})$, where $\mathbf{x}$ continues to appear only through the *linear predictor*, $\eta = \mathbf{x}'\boldsymbol{\beta}$ (see Equation 8). As a result, the mean of Y is determined by η and it is conventional to write $g(\mu) = \eta$, or $\mu = g^{-1}(\eta)$. This function $g(\cdot)$ is known as *link function*, as it links the mean, μ, and the linear predictor, η.

Thus, it is widely said that GLMs are a very suitable approach to deal with non-normality and non-constant models at the same time, and so the problem of the choice of designs is projected on this class of models too. According to what we mentioned before, GLMs can be completely characterized by their density function or by the likelihood function (Firth 1991). For example, let's consider a response variable that belongs to the exponential family of distributions, whose generic likelihood function is:

$$L(\theta, \phi; y) = \exp\left\{\frac{y\theta - c(\theta)}{\phi} + h(y, \phi)\right\} \tag{1}$$

for a single observation y, where $c(\cdot)$ and $h(\cdot)$ are taken as known functions. The quantities θ and ϕ are also known, and they are so called "natural" and "dispersion" parameters, respectively. Among some important properties GLMs have been related to the variance–covariance matrix of the response (McCullagh & Nelder 1989). The form of the variance of the observations for a given level of $\mathbf{x}$, varies as a function of the mean of the distribution that data come from:

$$V(y \mid \mathbf{x}_i) = \phi V(\mu) \tag{2}$$

where $\mu = E(y \mid \mathbf{x})$. The function $V(\mu)$ is the so called "variance function", which is conditional to $\mathbf{x}$ as well as μ. This function is associated with the specific distribution considered (*e.g.*, for the normal distribution, we have $V(\mu) = 1$ and $\phi = \sigma^2$). From Equation (2), we can see that $V(\mu)$ depends on the parameters of the model considered. Taking into account this aspect, iterative class of procedures have to be carried out on the way of estimating those parameters (*e.g.* Fisher scoring, Newton-Raphson, etc.). Thus, the estimation of $\boldsymbol{\beta}$ is usually expressed as:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{y}\right) \tag{3}$$

where $\mathbf{X}$ is the usual design matrix and $\widehat{\mathbf{W}}$ is the estimated diagonal matrix of "weights" (Myers et al. 2001a), given by $\widehat{\mathbf{W}} = \operatorname{diag}\{\widehat{w}_1, \ldots, \widehat{w}_n\}$, whose i-th element is given by:

$$\widehat{w}_i = \widehat{V}(y_i \mid \mathbf{x}_i)^{-1} \tag{4}$$

for $i = 1, \ldots, n$. Recalling Equation (2) and provided that $\mu = g^{-1}(\mathbf{x}'\boldsymbol{\beta})$, the dependence of the variance function on the parameters on the "weights" of the variance–covariance matrix of $\widehat{\mathbf{W}}$ can be seen:

$$\widehat{w}_i = \left(\phi V(\widehat{\mu}_i)\right)^{-1} = \left(\phi V\left(g^{-1}\left(\mathbf{x}'_i\widehat{\boldsymbol{\beta}}\right)\right)\right)^{-1} \tag{5}$$

From Equation (4), the dependence of the "weights", w_i, on the unknown parameters is an obstacle when choosing the "proper" matrix design $\mathbf{X}$ that leads to "good designs". This aspect is also clearly seen on the corresponding asymptotic variance–covariance matrix of $\boldsymbol{\beta}$, which is approximated (Khuri & Mukhopadhyay 2006) by:

$$\mathbf{V}\left(\widehat{\boldsymbol{\beta}}\right) \approx \frac{1}{\phi}\left(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\right)^{-1} \tag{6}$$

From Equation (6), an implicit "structural" consequence also follows: although the design matrix $\mathbf{X}$ could be orthogonal, the presence of the matrix $\widehat{\mathbf{W}}$ does not imply that $\mathbf{V}(\widehat{\boldsymbol{\beta}})$ should be diagonal. If we compare this situation with the linear case, this dependence on the unknown parameters conditions the choosing of optimal designs. In fact, considering a normal model, say $y \sim N(\mu, \sigma^2)$, and whose i-th observation can be represented as a linear function of its parameters, say $y_i = \mathbf{x}'_i\boldsymbol{\beta} + \varepsilon_i$, the quality of its variance–covariance matrix depends only on the choice of the design matrix, $\mathbf{X}$, that is: $\mathbf{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. When designing the experiment, orthogonality of $\mathbf{X}$ also gives some desirable statistical properties for the estimates of the parameters of the model. For a complete characterization of GLMs, see McCullagh & Nelder (1989), Dobson (2002) or Cordeiro & De Andrade Lima Neto (2004) among others.

1.3. Models for Binary Data

There are many processes –particularly in the industry– where a quality characteristic is modeled by a set of control factors. Processes with binary responses

can be modeled within the scope of GLMs, since Bernoulli or binomial distributions belongs to the exponential family. For example, we can consider a random variable with binomial distribution, say $y \sim \text{binomial}(m, \pi)$, where y is the number of "successes" out of m trials of a Bernoulli process, with a probability of success $\pi = P(y = 1)$. If the different values of π are put on the context of designed experiments, some experimental conditions can be considered as a function of a set of factors, represented by a k-dimensional vector of factors, say $x_1, \ldots, x_k$, whereas the i-th observed response, π_i, could be associated to the i-th experimental condition considered. Although we consider k known factors, there must be n different experimental conditions, $n > k$, to perform the experiment properly. If we use the logistic function to model the dependence relationship between π and the set of n different experimental conditions, this is usually stated as:

$$\pi(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})} \tag{7}$$

where $\mathbf{x}_i$ represents the i-th experimental condition considered and $\boldsymbol{\beta}$ is the usual column vector of unknown parameters, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$. From the usual theory, this model is clearly non-linear and its distribution cannot follow a normal form since it has an upper bound for proportions (*e.g.* 100%). Regarding the logistic model for the probability of success given on Equation (7), the usual way to define the linear predictor for the model is by defining a suitable link function. In this paper, we will use the *logit* link function, defined as $\ln\left(\frac{\pi}{1-\pi}\right)$, where $\pi = \pi(\mathbf{x}_i, \boldsymbol{\beta})$, as in Equation (7) (Collett 2002). Given this, it is easy to say that if we can apply the *logit* link function to (7), the resulting linear predictor will be $\eta = \mathbf{x}' \boldsymbol{\beta}$, where $\eta = \text{logit}(\pi)$. We can say that the response $\pi(\mathbf{x}, \boldsymbol{\beta})$ has been "linearized" by means of the link function. For example, if we consider $k = 2$ factors, the second order complete model for the linear predictor can be written as:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 \tag{8}$$

For a given design, *e.g.* given by its design matrix $\mathbf{X}$, and regarding Equation (6), each one of the elements of the matrix $\mathbf{W}$ are given by:

$$w_i = V(y_i | \mathbf{x_i})^{-1} = (m_i \pi_i (1 - \pi_i))^{-1}. \tag{9}$$

where m_i is the number of trials of the i-th experimental run, chosen from a Bernoulli process, with probability of success π_i, which is modeled as a function of $\mathbf{x}$ and $\boldsymbol{\beta}$, as in Equation (7).

From Equation (9), the probability of success depends on the values of the parameters. If a different link function is chosen (*e.g.* probit, complementary log-log, etc.), the situation is analogous (Myers et al. 2001*a*). Thus, it is clear that since the asymptotic variance-covariance matrix depends on the values of the parameters, optimal designs are difficult to achieve when abnormal responses.

2. The Choice of Designs

2.1. General Scope

The problem of what kind of designs, taken sequentially, may be suitable for a given process when an unknown relationship links the response to a set of factors has been studied deeply. When the errors are supposed to follow a normal distribution and linearity is presumably a good choice for the proposed model, optimal designs may also be obtained (Pukelsheim 1993). However, when other error structures arise in the data, the normal classical linear scope does not seem to be adequate (Collett 2002).

2.2. Looking for "good" Designs

Returning to the binary data family of problems, an experimenter may want to build empirical models for probabilities (*e.g.* for proportions) in terms of some control variables. For example, the experimenter may be interested in how to deal with the set up of the control factors so that the proportion of successes can be optimized (*e.g.* maximized). If a *specific* set up for these factors is found, we can link this situation with the one that is described by a single factorial design approach. We can call this situation a "static design", since a single design can show the experimenter what levels of factors are most suitable in order to optimize the proportion of good outcomes. Having this initial situation in mind, a basic question may be formulated: *"How could experimenters deal with this problem when the nature of the process does not follow the classical assumptions of normality and linearity?".* GLMs can offer an authoritative answer to this question. Classical factorial designs can be taken into account to find the best set up of the process to obtain the best outcome. We can say that the single-design solution for this situation has been studied and solved adequately. For example, Lewis et al. (2001*b*), Lewis et al. (2001*a*) and Myers et al. (2001*a*) describe the usual extension to factorial designs to solve this single-design solution.

At this point, we can divide this problem into a two–fold way. Both sides of the problem have special interest depending on the objective of experimentation. On one hand, the experimenter may be interested in finding the best design for a single experimental run. There are some approaches that we will summarize in the next section that consist of some contributions that some authors have made to clarify this problem. And on the other hand, the experimenter may want to set up specific strategies to sequentially explore a process, given a specific number of observations (*e.g.* fixed budget).

2.3. Optimal Single–Designs

An introduction on the problem of dealing with the choice of the "best" design is introduced by Khuri (1993, 2001). The author states the problem of the parameter dependence for GLMs when a sequential approach of optima conditions is followed.

A very good statistical scope is described here, especially from the point of view of the analysis of experiments.

In Atkinson (2006) a general formulation for optimal single–design problems is described in detail. Binomial models are also detailed for single and multiple variable cases. For the first case, the optimum allocation for a single variable design is explained from the D-optimum point of view. When the two-variables case ($p = 3$) is considered (*e.g.* with an associated linear predictor of the form $\eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2$), the author evaluates four sets of parameter values for the same experimental region. From the D-optimal point of view, the author discusses and explains point allocations for obtaining optimal designs. Although optimal configurations for a single design are explained here, the approach for sequential designs still remains unsolved.

Two articles, as Robinson & Khuri (2003) or Khuri & Mukhopadhyay (2006), propose a valuable graphical technique –the so called *quantile dispersion graphs* (QDGs)– whose details are exposed and analyzed. The first paper states some principles on the general formulation of the problem and two examples of the logistic model are examined as well. With relevance to the second one, an example using the Poisson distribution is also analyzed. The use of the mean-squared error of prediction (MSEP) as a basic criterion leads to the graphical approach of comparing and evaluating designs. Both papers focus on the same strategy: to minimize the MSEP($\mathbf{x}$) function over an experimental region, specially focused on the bias as a function of the mean, evaluated on different experimental runs, $\mathbf{x}_i$. As a result, these graphs provide an assessment of the overall prediction capability of a given design through a visual display of the MSEP.

In summary, the problem of finding optimal designs for a single experiment is being treated from more than one point of view, and its results are a very valuable contribution to further exploring the problem of dealing with the parameter–dependence problem.

2.4. A Viewpoint for Sequential Designs

In this article, we propose a scope to investigate different scenarios in the choice of a suitable strategy of exploration for an unknown process from the viewpoint of RSM. The specific problem that we present here is when the problem of sequential choice of designs is applied to processes whose main interest is modeling the proportion of success with a set of control factors. Since the associated distribution can be a binomial one, the toolkit that we use is the GLM approach. Thus, the next challenge to the experimenter may be expressed with other questions: *"How could one design a sequential strategy to learn about the operation of a system with binary response, when certain objectives are persecuted?"*, and more specifically: *"How could RSM and GLM be linked using the latter as a methodological support?"*. Taking into account the lack of precise and complete antecedents connecting RSM and GLM from a sequential experimentation point of view, we have been prompted to follow this form. We consider, as a particular strategy, the one in which the experimenter has a fixed number of observations to be made, in what is labeled as

"strategy of fixed budget". Thus, the objective will be to quantify the information gained once we have used all the budget available. In both cases, our plan is to carry out 2-level factorial and sequential designs.

3. A Proposal for Sequential Designs

3.1. Resharpening the Problem

We will turn our attention to processes in which a certain characteristic should be explained by a set of factors. For example, let's consider a typical Bernoulli process, with its only possible outcomes as $y_i = 1$ and $y_i = 0$. When associated to these outcomes, a binomial distribution may be linked with them in the following way: the probabilities of success and non-success can be stated as $P(y_i = 1) = \pi_i$ and $P(y_i = 0) = 1 - \pi_i$, respectively. When m_i identical trials of this process are drawn, a number of successes y_i can be counted, each one of them with the probability of success being equal to π_i.

Let's now put the projection of this problem into the context of classical experimental design. In order to do this, we consider three basic aspects, among others:

1. First: a number of experimental conditions, n, can be defined to run the experiment. If m_i trials of Bernoulli observations are considered for every experimental run, we can define a proportion of successes $p_i = \frac{y_i}{m_i}$, where y_i is the number of successes observed on the m_i trials considered. Thus, each experimental condition is a combination of levels of the set of factors considered.

2. Second: since the number of successes y_i that can be observed on each experimental condition is a random variable, then the associated proportion p_i is a random variable too, and

3. Third: since each value p_i is the observed proportion of m_i binary trials, and since its outcome depends on the experimental condition considered, the proportion itself cannot be controlled but it can be explained as a function of the control factors x_i.

Following classical experimental design, the next step is to define a proper model that can link the nature of the proportion as a response with the set of control factors as explicative variables. Since the greatest interest in the literature has been in logistic models for binary data (Atkinson 2006) we will use this model to link these variables. In these kinds of processes, a typical issue is to find out what levels of the factors considered lead achieving a goal for the response (*e.g.* maximize or minimize). Following the context of the problem described, we will follow the maximization objective for the proportion of successes.

3.2. Design Strategies

3.2.1. First Outline

In this article, we propose a *design strategy* like the one that consists of the choice of the levels of 3 variables, named w, L and S, that leads to a specific sequence of designs in order to select the "best" strategy of experimentation for a process. If $\boldsymbol{\psi}$ is a set of these 3 proposed variables, we can say that $D(\boldsymbol{\psi})$ defines the design strategy mentioned before. In the following sections, we present a proposed methodology based on the definition of two statistical criteria, named T_1 and T_2, both based on the amount of information on fitted models obtained after a fixed experimental budget was used. The evaluation of both criteria leads the experimenter to have an objective basis from which to decide what is the best alternative to perform a RSM–based exploration of a process in which a proportion is used as a response. Thus, we will examine a typical sequence of experimental designs whose quality is examined by means of the evaluation of these two criteria.

Although the response variable is a proportion, we follow a classical RSM–based methodology, which is the usual way to attain optima conditions. Some extensions of it are exploited here for logistic models. Response surfaces for binomial models are evaluated from a two–way viewpoint: (a) on one hand, the GLM approach to deal with non-normal models, in particular for those that are modeled by logistic regression ones, and (b) on the other one, experimental design scope is focused here in its classical way, using two-level factorials designs and adapting the classical methods of RSM. The choice of the levels of these 3 evaluation variables defines different strategies to evaluate experimental designs. In summary, for a given scenario of evaluation variables, the goal will be to find the best design strategy in order to achieve the maximum value of the binomial response, when a set of two factors are defined to perform this exploration. In next sections, we progressively describe all the elements this strategy.

3.2.2. Main Components of the Strategy

The approach that we propose here starts with the definition of the following items:

1. *Response*: all the observations come from binomial processes as proportions, p_i (see Tables 1, 2 and 3). For simplicity, the number of trials of Bernoulli observations has been fixed as $m_i = 100$ for each experimental condition. As stated before, the variable of interest is the probability of success, explained by second order models.

2. *Factors*: we consider complete second-order linear predictors of 2 factors, x_1 and x_2, each of them at 2 levels and fixed.

3. *Factorial designs*: for the observation of the proportions at different factor levels, factorial designs at 2 levels have been considered.

4. *Theoretical surface*: the process that generates the data is represented by a theoretical surface, whose form has been simulated computationally in R (R Development Core Team 2006) by considering complete second order logistic models. We will refer to this surface as $\pi(\mathbf{x}, \boldsymbol{\beta})$, indicating that there is an explicit dependence on the factors vector, $\mathbf{x}$, and on the unknown parameters vectors, $\boldsymbol{\beta}$, as well.

5. *Fixed budget*: this is an outer constraint that we have fixed in order to compare different strategies of exploration and to find the best one. This number has been stated in $n = 1500$ binary observations, which is the one that the experimenter has at the beginning of the experimentation.

The first consideration about our proposal brings the experimenter into the following situation: *how to gain the best information quality of the design of an experiment having a fixed budget to experiment.* As a starting point, we have considered the fixed budget as the one that consists of 15 experimental points available. We assign a scheme of 3 successive two–level factorials, each of them consisting of 5 design points: four on the vertices of the region defined by the two levels of the factors and the last design point assigned to the center of the factorial. For all of these experimental conditions, a number of proportions of successes are observed. Following a sequential definition of successive designs –driven by classical RSM approaches of exploration, *e.g* steepest path– logistic models are also fitted to data. At end of this exploration (*e.g.* when all the budget has been used and when a final model has been fitted) two evaluation criteria are also defined in order to measure the quality of the designs, which leads the experimenter to achieve the maximum conditions for the response. Figure 1 shows the evolution of the sequential factorials, exploring the surface from the first design (with center O_1) to the third and last one (with center O_3). The fitted model is represented here by its contour levels, each one of the form $w \cdot \pi_{\max} = \text{constant}$, whose details will be described in the next section.

3.2.3. Evaluation Variables

As stated before, we have defined 3 variables to perform the evaluation of different design strategies. We will consider each variable at 5 levels, which we describe as follows.

- The "center variable", w: the definition of this variable aims to define the experimental center point for the first design. See Figure 2. If the maximum value for $\pi(\mathbf{x}; \boldsymbol{\beta})$ is $\pi_{\max}$, then w is defined as a fraction of this maximum value. When $w = 0$ and when $w = 1$ the surface $\pi(\mathbf{x}; \boldsymbol{\beta})$ will be bounded between two planes, respectively: the factors plane ($\pi = 0$) and the top bound plane ($\pi = \pi_{\max}$). Then, w defines an intermediate plane between these two bounds, so that the theoretical surface is cut with a parallel plane to the factors one at the height $w \cdot \pi(\mathbf{x}; \boldsymbol{\beta})$. This intersection defines a contour line that is projected on the factors plane of the form $\pi(\mathbf{x}; \boldsymbol{\beta}) = \text{constant}$, in which the first experimental center will be located.
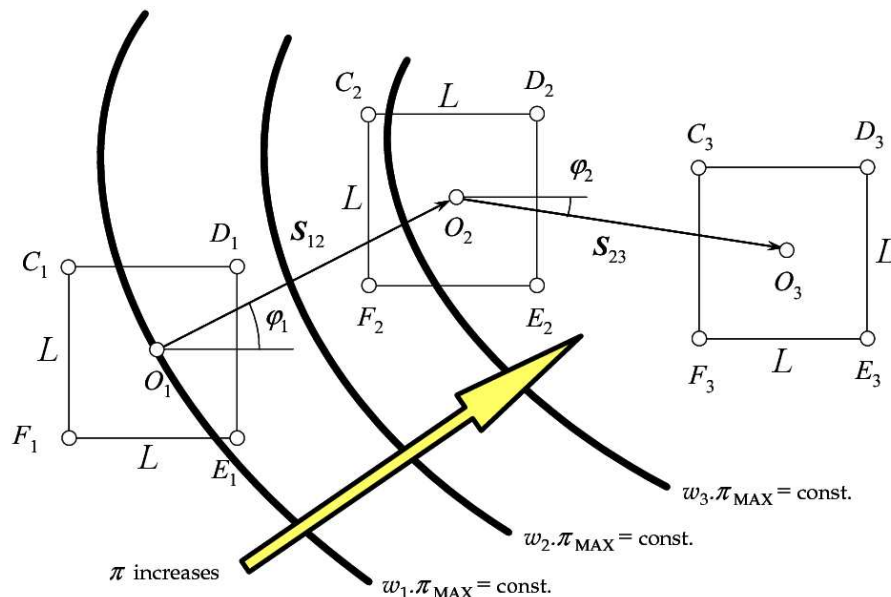
FIGURE 1: Evolution of sequential designs for a fixed value of w, L and S. The main arrow represents the sequential approximation to maxima conditions of the response, as in classical RSM approach.

- The "side variable", L: this variable refers to the length of the "side" of a square of side L. Factorials will represent sequential experimental regions on the factors' plane, each one of them with usual limits, (-1) and $(+1)$ for each side of length L, expressed in coded units (see Figure 1).

- The "leap variable", S: this variable is the modulus of the vector that separates two experimental centers one from the other, the same way as in classical RSM. The notation S_{ij} refers to the leap vector that connects centers O_i and O_j, whose direction is given by ϕ_i.

Figure 1 shows how do L and S look.

Having defined a set of 5 levels for the 3 evaluation variables, the main objective will be to determine which are the ones that lead to the maximum probability of success for a given strategy. If $\psi = (w, L, S)'$ is the vector that summarizes these 3 evaluation variables, then the design strategy is defined by $D(\psi) = D(w, L, S)$. Thus, the goal will be to choose values for ψ to achieve the maximum value for $\pi(\mathbf{x}, \boldsymbol{\beta})$. This value of ψ will define the "best" design strategy, $D^*(\psi) = D^*(w, L, S)$, which will be objectively evaluated from the point of view of T_1 and T_2, as we will see next.
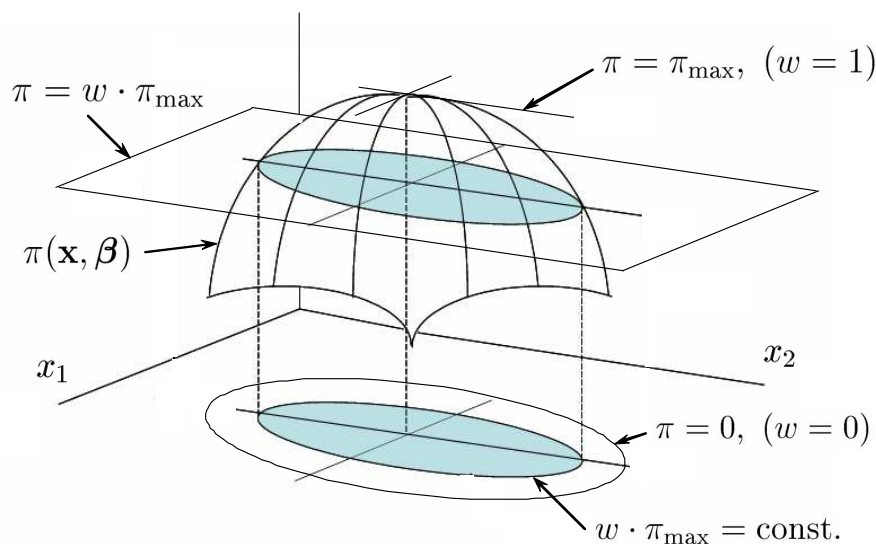
FIGURE 2: Generating a contour to place the first center point from w.

3.3. Description of the Strategy

On the study of this proposed design strategy in terms of $\boldsymbol{\psi}$, we will describe now 4 consecutive stages in order to facilitate a more complete development and evaluation of the problem. As a result, the last stage (number IV, as we will see immediately) defines the problem completely. This point of view basically consists of the random generation of observed values for the response on the design points, for which logistic models are fitted. In this sense, the idea is to evaluate different strategies based on different configurations for ψ. After this, all strategies will be measured and compared with each other, choosing the one that best satisfies the two criteria proposed to measure the quality of the associated designs.

1. *Stage I: one case study, all variables fixed.* The starting situation lies in a fixed budget of $n = 15$ design points. Observations of a binomial response –with $m = 100$ trials each– are generated at each design point. These 15 design points are arranged on 3 successive 2^2 factorial designs of 5 design points each: 4 design points for vertices and a center point on every one. These 5 design points and these two factors considered may be allowed to fit a 5-parameter model, *e.g.* a model with 2 first-order terms, 1 for the interaction and 1 for the second-order term. More than 1 central point could also be taken in order to have a potential measurement of curvature, (*e.g.* a complete second-order model with $p = 6$ parameters). In Figure 2, points labeled with C_1, D_1, E_1 and F_1 illustrates these 4 design points for the vertices of the first factorial, whereas O_1 indicates the center point for the first design, as well. This set of $15 \times 100 = 1500$ points will be identified as a "case study". If we follow the classic RSM approach, a first model of 5 parameters can be fitted for these 5 design points by means of the GLM

approach. To do this, we use the logit link. The resulting estimates define the direction of the steepest path in the same way as in classical RSM. The first leap, $\mathbf{S_{12}}$, defines the second experimental center, around which the second factorial is set up. For fitting the second model, we considered both first and second design points. Thus, having $5 + 5 = 10$ design points, a complete second order model, whose estimates define the direction of the second leap, $\mathbf{S_{23}}$, is fitted. Having this point as the third experimental center, these new 5 design points leads to the use of the whole the budget available. In this situation, a final model is fitted with all the information available, that is, for all the 15 design points. For all this model fitting, a fixed level for $\boldsymbol{\psi}$ –that is, single fixed values for w, L and S– has been considered.

2. *Stage II: fifteen cases study, all variables fixed.* This second stage consists of a first generalization of previous Stage I. That is, starting with the same experimental center point and using the same levels for the evaluation variables, we have ran 15 times the generation of design points, using the same values of w, L and S. As a result, 15 cases study were evaluated with respect to fitting 15 corresponding models for each one. The sequential exploration of the theoretical surface and the model fitting are both analogous to the procedure of Stage I. Having all the information of the final fitted models for each case study, we propose two criteria of quality fitting: (a) Criterion I: the maximization of a Fisher Information Matrix determinant–based criterion for the last fitted model, and (b) Criterion II: the maximization of the information related to the location of the point of the fitted model that maximizes the theoretical model. Both are described as follows.

a) *First criterion*: The resultant fitted model for every one of the 15 case studies is of the form of expression (7), with the estimated values of $\boldsymbol{\beta}$ instead of the original ones, that is:

$$\widehat{\pi}(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}) = \frac{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})} \tag{10}$$

The variance–covariance matrix of these estimates, $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$, provides a first measurement of how well the model fits for a single case study. Provided that all the estimates of the parameter vector are of the inverse of $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$ is asymptotically equal to the the Fisher Information Matrix (abbreviated from here as "FIM") of this model. That is $\mathbf{I}_F = \mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$, where $\widehat{\mathbf{W}}$ is the estimated matrix of "weights", whose i-th element is given by Equation (9), being $\widehat{\pi}_i = \widehat{\pi}(\mathbf{x}_i, \widehat{\boldsymbol{\beta}})$, as in expression (10). Taking the determinant of this FIM, this measurement is concentrated in only one figure. A logarithmical transformation of this determinant is taken next, so the quantity $\log \det(\mathbf{I}_F)$ is finally taken as a proposed measurement for each one of the 15 case studies generated for a single level of w, L and S. As a global measurement of all these 15 runs –that is, for the 15 case studies– we take the average of all of them which leads to another unique measurement of how good it fits. We call this

the T_1 statistic, which will be used as the first criterion of evaluation of fitted models. For a general number of cases study, say G, this statistic is defined as:

$$T_1 = \frac{1}{G} \sum_{i=1}^{G} \log \det(\mathbf{I}_F)_i, \qquad 0 \le T_1 \le +\infty \qquad (11)$$

The motivation of defining T_1 as in (11) is to have a representative measurement (*e.g.* the average) of how much information is contained on a single design strategy $D(\psi)$. Since every value of the latter defines an index of the quality of design strategy, it will be relatively easy to compare different values of w, L and S for different explorations of the process, and thus, to look for the "best" design strategy, $D^*(w, L, S)$, if there was any. As mentioned before, we will take $G = 15$.

b) *Second criterion*: This second measurement starts with the "best" conditions that the fitted model can give for a certain chosen strategy. In particular, and given the fitted model by (10), the fitted surface will reach a maximum when the factors have a certain value $\widehat{\mathbf{x}}_{\max}$, so that $\frac{\partial \widehat{\pi}(\mathbf{x}, \widehat{\boldsymbol{\beta}})}{\partial \mathbf{x}} = 0$, if $\mathbf{x} = \widehat{\mathbf{x}}_{\max}$. Evaluating this value $\mathbf{x} = \widehat{\mathbf{x}}_{\max}$ on the true surface given by $\pi(\mathbf{x}, \boldsymbol{\beta})$, a comparison with the "best" operative conditions can be made. That is, provided that $\pi(\mathbf{x}_{\max}, \boldsymbol{\beta})$ leads to the best approximation to the maxima conditions in the theoretical model, the best experimental conditions obtained from a fitted model will be $\widehat{\pi}(\widehat{\mathbf{x}}_{\max}, \widehat{\boldsymbol{\beta}})$. For example, Figure 3 shows all these measures for the single factor problem. In this Figure, two models can be identified: on the left hand, the true one, labeled with $\pi(\mathbf{x}, \boldsymbol{\beta})$, and on the right hand, the fitted one, $\widehat{\pi}(\mathbf{x}, \widehat{\boldsymbol{\beta}})$. The best theoretical condition is given by the maximum point, $\pi(\mathbf{x}_{\max}, \boldsymbol{\beta})$, whereas the best one obtained for the fitted model is represented by $\widehat{\pi}(\widehat{\mathbf{x}}_{\max}, \widehat{\boldsymbol{\beta}})$. Evaluating this fitted maximum condition $\widehat{\mathbf{x}}_{\max}$ on the theoretical model, it results in the value $\pi(\widehat{\mathbf{x}}_{\max}, \boldsymbol{\beta})$. As the point $\widehat{\mathbf{x}}_{\max}$ moves towards the true maximum point $\mathbf{x}_{\max}$, the maximum response also tends to $\pi(\mathbf{x}_{\max}, \boldsymbol{\beta})$, which is the maximum theoretical value that the response can achieve. If $\pi_i(\widehat{\mathbf{x}}, \boldsymbol{\beta})$ characterizes the i-th case study, we can summarize all the 15 cases study considered by means of taking the average of them. For a general number of G case studies, we define this T_2 statistic as:

$$T_2 = \frac{1}{G} \sum_{i=1}^{G} \pi_i(\widehat{\mathbf{x}}, \boldsymbol{\beta}), \quad 0 \le T_2 \le 1 \qquad (12)$$

As it was explained for T_1, this T_2 statistic follows an analogous motivation too, now for the "best" experimental condition obtained by the process of model fitting, that is $\widehat{\mathbf{x}}$. To perform a coherent comparison between both criteria, T_2 was thought of as a function related also to ψ in order to compare it with T_1. This comparison is then performed by looking at which levels of these variables w, L and S are the ones that

maximize each criterion. So, using this result T_2, the second proposed criterion to evaluate the quality of the design strategy will be examined next.
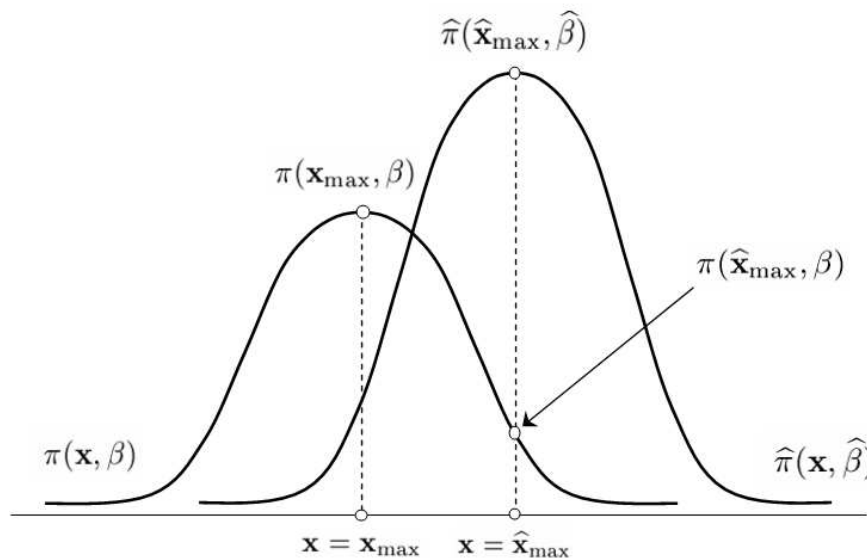

FIGURE 3: General scheme for the second criterion, T_2, for the single factor problem.

3. *Stage III: fifteen case studies, w fixed and L and S variables.* In this next stage, 5 levels are now considered for both L and S, keeping w fixed. This extension leads us take into consideration now a scenario of 25 cases, each one developed the same way as in the previous stages. These 25 generalizations are made in the same way as the ones described in Stage II. This scheme is summarized in a 25×2 matrix, that is called the "**LS** matrix". Each row of it defines one of the 25 cases considered, and for each one, 15 runs are evaluated. As a result of this generalization, both measurements are obtained as in the previous Stage. Either T_1 and T_2 are calculated for every one of the 25 cases considered here.

4. *Stage IV: fifteen case studies, with w, L and S all variables.* This final stage aims to end the generalization started in the previous ones. Five levels of w are now considered here, so that for every row of the **LS** matrix defined in Stage III, there are 5 different experimental conditions, each one given for all the 5 levels of w considered. Having all the $5 \times 25 = 125$ cases, we calculate T_1 and T_2 for each one.

As a result of this 4 stage–development, we have now summarized the situation as follows: (a) We started with 5 frames, each one of them characterized by a single level for w and 5 levels for L and S, respectively. Thus, each frame consists of 25 points on a 5×5 grid for L and S. (b) Values of T_1 and T_2 are calculated for every

frame. Then, 5 grid frames –each one for every level of w considered– have been obtained, that consist of 5×5 values for either T_1 and T_2. Finally, and since we are evaluating all the strategies considered in terms of vector $\boldsymbol{\psi}$, we have taken as "best" strategies the ones that maximizes both T_1 and T_2. In the next section we illustrate the procedure for a worked example.

4. Examples

4.1. Setting Up the Problem

A simulated process of random generation of data is considered here. The idea with this is to study in detail the behavior of a real process characterized by Bernoulli observations. Then, an RSM approach will be developed following the strategy we have mentioned before. We will describe now the four stages mentioned in previous section for a simulated example obtained with R. These stages are developed here in order to measure the quality of designs for various strategies. This next example is explained for a single value of the evaluation variables. An extension of it will be considered later, including all the levels of w, L and S, as well.

4.1.1. The True Surface

Two factors, x_1 and x_2, have been considered when generating the true process surface. It is then described by means of a second order model for these two factors, whose linear predictor is of the form given by expression (8). The levels of each one of the fixed constants β_j of the generation process, $j = 1, \ldots, 6$, were chosen randomly. Thus, the resulting value of this column vector was $\boldsymbol{\beta} = (2, 2, -2, -2, -2, 2)'$, which leads to the expression of the theoretical surface, expressed by the logistic model given by expression (13), and whose graphical form is represented in Figure 4:

$$\pi(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp(2 + 2x_1 - 2x_2 - 2x_1^2 - 2x_2^2 + 2x_1 x_2)}{1 + \exp(2 + 2x_1 - 2x_2 - 2x_1^2 - 2x_2^2 + 2x_1 x_2)} \tag{13}$$

In order to obtain only configurations of this vector that correspond to maxima situations, all the models obtained were filtered according to the criteria of eigenvalues classification. Since second order polynomials were considered for the linear predictor, we retained those where both eigenvalues had a negative sign. Thus, the complete characterization of this surface can be summarized by: (a) Eigenvalues: $\lambda_1 = -1.0$ and $\lambda_2 = -3.0$. Both negative signs confirm that the surface has a maximum. (b) Coordinates of the stationary point: $\mathbf{x}_{\max} = (x_{1S}, x_{2S})' = (0.3333, -0.3333)'$. (c) Value of the response at $\mathbf{x}_{\max} : \pi(\mathbf{x}_{\max}, \beta) = 0.93504$. The surface is given, therefore, by Equation (13), and it represents a true process from which all the data of the experiment (*e.g.*, proportions) is obtained. Since its values are bounded between 0 and 1, this surface can be thought of as a generator of proportions as responses (*e.g.* y is the number of successes obtained from a binomial
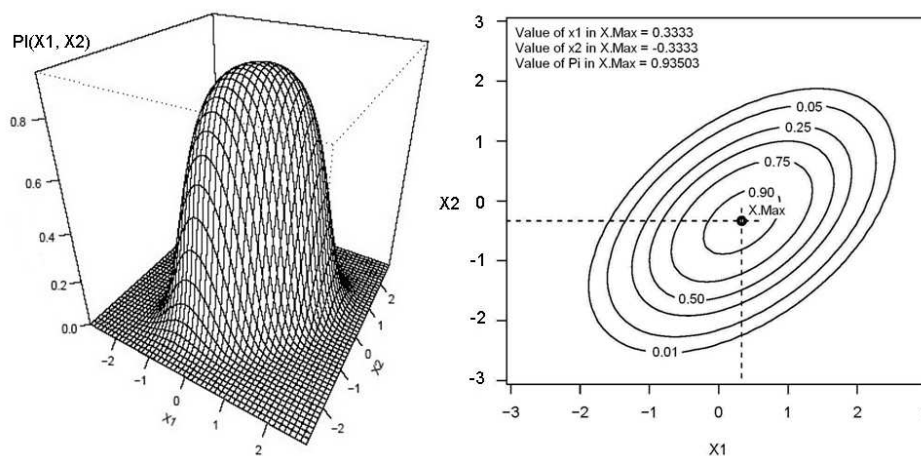
FIGURE 4: The theoretical surface, representative of the process studied.

distribution indexed by its size m and its probability of success π). Putting this response variable in the context of the design of experiments, it can be modeled using the logistic function as $\pi(x_1, x_2, \boldsymbol{\beta}) = \pi(\mathbf{x}, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the 6×1 vector of unknown parameters that appear in (13). Following this way, then the surface represents how π varies on the design points of the factors plane. Specifically, the shape of this figure is associated with the form of the logistic model previously defined in Equation (13) and Figure 4. The objective will be, then, to select different sequential designs that lead to the attainment of maxima conditions for $\pi(\mathbf{x}, \boldsymbol{\beta})$.

4.1.2. The First Design Point

Having defined a way to represent the real process by the theoretical surface, the next step is to start exploring it from the point of view of sequential experimental designs. It is clear that the nonlinear nature of the real model –see expression (13)– forces the experimenter to perform a *sequential* exploration of the true process. This is because he or she knows neither how this process works, nor the form in which the factors explain the response variable. A one-shot experiment, in which the whole budget was used, is definitely not a wise way to proceed, especially at the beginning of the experimentation (see Box et al. (2005), for example). We develop the exploration of this surface by the use of factorial designs at two levels. We will describe here how we adapted this methodology to the problem.

The first step is to define a way to set up a design point into an operability region of the true process. In order to do it systematically, we localize this point over a certain contour level of the theoretical surface. Figure 2 illustrates this situation. A particular contour line arises when a certain value for the variable w defines a plane that intersects the surface in a parallel way to the factor's plane. $\pi_{\max}$ being the maximum value that the surface can take, then the equation of this horizontal plane can be expressed as $w \cdot \pi_{\max} = $ constant, which is also an

analytical form of the contour line projected on the factor's plane. As a starting point, we have set up the value of w in 5%. This situation can be also illustrated in Figure 2. The corresponding contour level will be, then, the locus in which the first design point will be set up. In order to define another way to generate a first design point systematically, we connected the origin of the factor's plane, $\mathbf{x}_0 = (x_{01}; x_{02})' = (0; 0)'$, with the coordinates of the maximum point of the surface, $\mathbf{x}_{\max} = (x_{1\max}; x_{2\max})'$. As a result, these points define a segment line R_1 that intersects the contour level in two symmetric points around $\mathbf{x}_{\max}$.

4.1.3. Levels for Evaluation Variables

As stated before, a value for the "center variable", w, is needed for the definition of the contour level that will contain the center point for the first design. For example (see Figure 2), small values for w will lead to the definition of bigger contour lines and thus small values for $\pi(\mathbf{x}, \boldsymbol{\beta})$. In the same way, the neighborhood of the maximum will also be around large values of w, since the studied case is a maximum. Because it is reasonable that the experiment probably starts on regions far away from the optimum conditions (Box & Draper 1987), it can be also reasonable to start experimenting on regions in which the values for the response are "small" too. Connecting this situation with w, relatively small values for this "center variable" are considered as a real representation for the beginning of the experiment. For this situation, then, we have defined the following levels for $w : w_1 = 5\%, w_2 = 10\%, w_3 = 15\%, w_4 = 20\%$ and $w_5 = 25\%$, because they can reasonably represent the situation at the beginning of a typical experimentation, especially here, since the studied case is a maximum.

On the left side of Figure 5 it can be seen how the point O_1 has been chosen as the first design point, which is over the contour line, labeled on the right figure as "CL at w = 0.05".
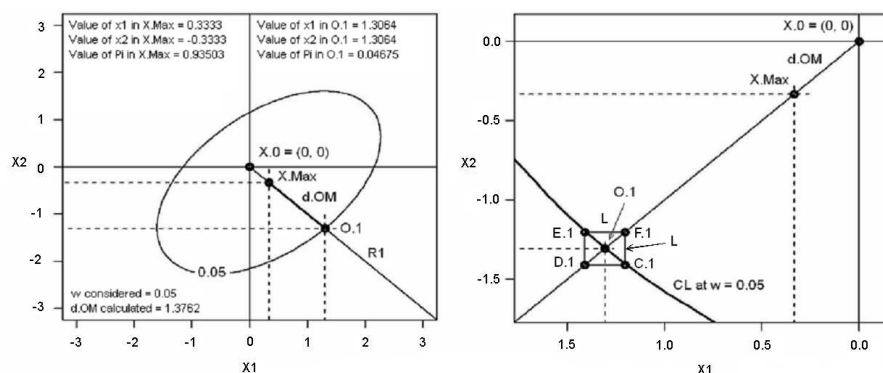

FIGURE 5: Left: the determination of the first experimental point, O.1 for a 5% contour level. Right: the first factorial design at two levels, with side L and with O.1 as its center. (The graphical notation O.1 refers to O_1, C.1 to C_1, etc.)

In the same figure, the distance between both points O_1 and X.Max, labeled as d.OM, was also automatically calculated with R and it is equal to 1.3762. For the determination of the levels for the second evaluation variable, L, we take the fourth part of d.OM and define it as an initial basic distance. We define 5 levels for L as a fraction of this distance. These are of the form $l \times 0.25$ d.OM, where l represents the fraction mentioned before. Large values for l may lead to make a relatively wide mapping of the factors region, while small ones will map small regions. In order to have a relatively informative range of variation, we have considered the following levels for $L : L_1 = 0.6 \times 0.25$ d.OM, $L_2 = 0.8 \times 0.25$ d.OM, $L_3 = 1.0 \times 0.25$ d.OM, $L_4 = 1.2 \times 0.25$ d.OM and $L_5 = 1.4 \times 0.25$ d.OM.

For the selection of levels of S the same logic is followed: small values for it may lead the experimenter to consider relatively many leaps to reach the maxima conditions, while large ones, few steps. In this sense, we evaluate the following levels for $S : S_1 = 0.2 \times$ d.OM, $S_2 = 0.4 \times$ d.OM, $S_3 = 0.6 \times$ d.OM, $S_4 = 0.8 \times$ d.OM and $S_5 = 1.0 \times$ d.OM.

In summary, for a given process (see Equation 13) a fixed value for w determines the position of the first experimental center. Once this point is set up, there is an associated distance, d.OM that leads to set up the other two evaluation variables, L and S. For the first one, each level is taken as multiples of the fourth part of d.OM, whose "weights" are: $0.60, 0.80, 1.00, 1.20$ and 1.40. And for the last one, S, its associated levels are taken directly as submultiple of d.OM, with the following "weights": $0.20, 0.40, 0.60, 0.80$ and 1.00. Therefore, $w = 0.05$ determines d.OM $= 1.37624$, and thus all levels for L and S are taken as the coordinates of respective vectors, as defined before:

$$\mathbf{L} = (L_1, L_2, L_3, L_4, L_5)' = (0.20644, 0.27525, 0.34406, 0.41287, 0.48168)' \quad (14)$$
$$\mathbf{S} = (S_1, S_2, S_3, S_4, S_5)' = (0.27525, 0.55049, 0.82574, 1.10099, 1.37624)' \quad (15)$$

4.2. An Example Considering a Single Row of LS(w)

4.2.1. A Single Experimental Run

Considering both vectors $\mathbf{L}$ and $\mathbf{S}$, we define the $\mathbf{LS}$ matrix, $\dim(\mathbf{LS}) = 25 \times 2$, which depends on the level of w considered. This matrix is the one that contains all the levels for both variables L and S, and that will be evaluated in different design strategies. Since five levels for L and S are considered, every row of this $\mathbf{LS}$ matrix is of the form $L_i S_i$, for $i = 1, \ldots, 5$. Besides which, and since five levels of w are considered, the complete scope of experimentation considers five matrices $\mathbf{LS}$, named: $\mathbf{LS}(0.05)$, $\mathbf{LS}(0.10)$, $\mathbf{LS}(0.15)$, $\mathbf{LS}(0.20)$ and $\mathbf{LS}(0.25)$.

For example, if we consider the true surface defined as in (13), and if $w = 0.05$, then the first design point, 0.1, is automatically located on the $0.05 \cdot \pi(\mathbf{x}_{\max}, \boldsymbol{\beta}) =$ const. contour level. The calculated distance from this point to the maximum is d.OM $= 1.3762$. Figure 5 shows this situation for this specific value of w. Immediately, the associated levels for $\mathbf{L}$ and $\mathbf{S}$ are the same as in expressions (14) and (15), respectively. Therefore, arranging both vectors in terms of its levels, we

obtain the **LS**(0.05) matrix, consisting of 25 rows of the different combinations of levels for L and S that we defined previously.

Now we can arrange all the aspects described before in order to build the first factorial as a beginning of the evaluation of different design strategies for $\pi(\mathbf{x}, \boldsymbol{\beta})$. We define now the first factorial design, which consists of 5 design points for both factors x_1 and x_2 (see the right side of Figure 5). This first factorial consists, then, of 4 vertices –that we can call: `C.1`, `D.1`, `E.1` and `F.1`– and a central point, `O.1`. For these 5 design points, the value of the "true" response is determined using the "true" model defined in (13). The layout of this first design with the levels for both factors is represented on the first four columns of Table 1 for $L = 0.2064$.

TABLE 1: The layout for the first design, which contains the true response, $\pi(\mathbf{x}, \boldsymbol{\beta})$ for every design point. It also contains the observed responses, y, or the number of successes on a fixed size M, or pr, and the observed proportion of successes.

Design Point	x_1	x_2	$\pi(\mathbf{x}, \boldsymbol{\beta})$	M	y	pr
C.1	1.20323	-1.40966	0.04581	100	5	0.05
D.1	1.40966	-1.40966	0.01360	100	2	0.02
E.1	1.40966	-1.20323	0.04581	100	6	0.06
F.1	1.20323	-1.20323	0.13312	100	13	0.13
O.1	1.30645	-1.30645	0.04675	100	7	0.07

In order not to deal with theoretical data, but with real ones, we perform a random generation process for the responses, using the true response as the expected value of a binomial process. This is indexed by m and $\pi(\mathbf{x}, \boldsymbol{\beta})$, $\mathbf{x}$ being every one of the design points given on the first three columns of table 1. Regarding the nature of the process being studied, all these generated responses can be assumed to be observations from a binomial process, whose probabilities of success are the dependant variable on the 2 factors-model described in Equation (13). Therefore, a random generation of data of the form $y_i \sim \text{binomial}(m_i, \pi_i)$ can be stated this way, in which y_i are the number of successes observed on m_i trials of a binomial process with probability $\pi_i = \pi(x_{i1}, x_{i2}; \boldsymbol{\beta})$, for each one of the design points described before. Figure 6 illustrates a scheme of the aspect of generated observations of the response based on the expected value for a generical point $\mathbf{x}_0$. Having the expected value of the true response at that point, say $\pi(\mathbf{x}_0, \boldsymbol{\beta})$, and using the R command `rbinom(n,size,prob)`, this notation refers to `n` as the number of observations that will be generated; `size` is the number of trials considered and `prob` represents the probability of success for each trial, chosen for the generation of data. This probability is the one that is modeled with the logistic model given in Equation (13). The same figure shows how random values were generated around this last one.

On the definition of the proportions as the responses, we have fixed the "index" (McCullagh & Nelder 1989), m_i, of the binomial distribution as 100. So, this situation is equivalent to the observation of 100 single trials of a Bernoulli process with probability of success begin equal to the one described in Equation (13). The fifth and sixth columns of Table 1 reflect the results of this situation, so the sixth one is the number of observed "successes" in every experimental condition following the generation process described in this way. Finally, dividing the column of
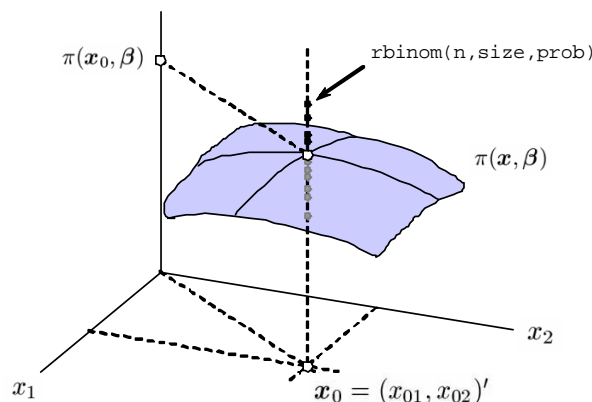
FIGURE 6: The random process of generating data around expected values.

observed successes `y` by the index `M`, it results into a vector of observed proportions `pr` for each one of all the 5 experimental conditions defined.

All these 5 points of Table 1 –labeled as column "`pr`" on it– can be fitted with a suitable model for proportions, such as the logistic model described in (7).

Since we have that $\mathtt{pr}_i = \widehat{\pi}(\mathbf{x}_i, \boldsymbol{\beta})$ are the observed proportions for every design point, the logistic model for this first design is of the form $\mathtt{pr}_i = \frac{\exp(\eta_i)}{\exp(1+\eta_i)}$, where $\eta_i = 2+2x_{i1}-2x_{i2}-2x_{i1}^2-2x_{i2}^2+2x_{i1}x_{i2}$ is the linear predictor, as described before. Provided that we have 5 points for the first design, this linear predictor has to be defined with a maximum of $p = 5$ parameters. Thus, for every design point, the response variable $\mathtt{pr}_i$ is expressed in terms of the linear predictor for the two factors considered, x_1 and x_2, as in Equation (13). The nature of this Equation leads to the performance of likelihood-based methods for the estimation of the unknown parameters β_j. As we mentioned before, GLMs uses this approach. In this article we use the logit link for the observed proportions. So, the model can be expressed as a function of the linear predictor. Using the `glm()` function of R, the program calculates the ANOVA table. As a result, and since we have $p = 5$, the program gives the maximum-likelihood estimates for $\boldsymbol{\beta}$, that is, $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_{12}, \widehat{\beta}_{11})'$, in which its standard errors and p-values are also obtained by R. Replacing these estimates on the true one, the fitted model can be expressed as:

$$\mathtt{pr} = \widehat{\pi}\big(\mathbf{x}, \widehat{\boldsymbol{\beta}}\big) = \frac{\exp\big(\widehat{\eta}\big)}{1 + \exp\big(\widehat{\eta}\big)} \tag{16}$$

The estimated vector $\widehat{\boldsymbol{\beta}}$ provides a way to find the direction of steepest path for the fitted model. Deriving from Equation (16) with respect to both factors, the direction of the leap vector $\mathbf{S}$ –which is, the third evaluation variable– can be determined as:

$$\mathbf{S} = (S_{12}\cos\varphi_1; S_{12}\sin\varphi_1)' = (s_1; s_2)' \tag{17}$$

whose direction, φ_1, is calculated in the same way as in classical RSM, that is $\varphi_1 = \arctan\frac{s_2}{s_1}$. In this example, we choose the first level of S –defined before–

for the modulus of this vector, so it will be $S_{12} = 0.2752474$. The direction of this vector is also calculated and is given by $\varphi_1 = 95.27759°$, so its components are: $s_1 = -0.02531755$ and $s_2 = 0.27408057$.

Now, if $\mathbf{P}_1$ represents any of the 5 coordinates of the points of the first factorial, the first leap vector –which is denoted by $\mathbf{S_{12}}$– can be added vectorially to them, so that the second design can be obtained as a sum of the form $\mathbf{P}_2 = \mathbf{P}_1 + \mathbf{S}_{12}$, where $\mathbf{P}_2$ corresponds to the coordinates of the second design points. Table 2 illustrates this new design.

TABLE 2: The layout and main results of both first design (points C.1, D.1, E.1, F.1, O.1) and second one (points C.2, D.2, E.2, F.2, O.2), both arranged as two sequential designs.

Design Point	x_1	x_2	$\pi(\mathbf{x}, \boldsymbol{\beta})$	M	y	pr
C.1	1.20323	−1.40966	0.04581	100	5	0.05
D.1	1.40966	−1.40966	0.01360	100	2	0.02
E.1	1.40966	−1.20323	0.04581	100	6	0.06
F.1	1.20323	−1.20323	0.13312	100	13	0.13
O.1	1.30645	−1.30645	0.04675	100	7	0.07
C.2	1.17791	−1.13558	0.19745	100	11	0.11
D.2	1.38435	−1.13558	0.07473	100	6	0.06
E.2	1.38435	−0.92915	0.18168	100	19	0.19
F.2	1.17791	−0.92915	0.38311	100	38	0.38
O.2	1.28113	−1.03237	0.19273	100	27	0.27

For this new second design, we proceed in the same way as in the first one. Keeping the same true model –given by (13)– and the same index for the binomial trials, we have to generate $m_i = 100$ Bernoulli observations, each one with probability of success given by (13), whose design points, $\mathbf{x}_i$, are the ones from the second factorial. The number of observed successes will be represented by y_i and the associated proportion of them will be $pr_i = \frac{y_i}{100}$, for $i = 1, \ldots, 5$. As a result, an extended table of 10 rows is also obtained, following the same structure as Table 1 but with 5 new rows, that correspond with points labeled as C.2, D.2, E.2, F.2, with O.2 as the second center point. All this is shown on Table 2.

The second fitted model now will contain all the information of the true response, provided by both designs. The way of fitting this second model is analogous to the first one, except now there are 10 degrees of freedom, which come from both designs. A complete second-order linear predictor of $p = 6$ parameters can now be considered and the way of fitting it follows the same logic as in the first design. Estimations of the parameters and also their standard errors and p-values are also obtained using the glm() R command in the same way as before. Now having the estimates $\widehat{\beta}_j, j = 1, \ldots, 6$ for this second model, the second leap, $\mathbf{S}_{23}$ can be determined the same way as in the first factorial. This second leap is then characterized by its angle $\varphi_2 = 95.8756°$ and by its components, $s_1 = -0.0280901$ and $s_2 = 0.2738103$.

Following all these steps for the third factorial, the third design is obtained. At this time, all the available budget has been used. That is, the three defined designs contain the information given by 1500 experimental points, each one of

5×100 Bernoulli observations. Thus, we set up the analysis fitting a final model for the three models, so that it contains all the information of the 1500 points. The linear predictor for it has $p = 6$ parameters, and a final vector of estimates $\widehat{\boldsymbol{\beta}}$ is also obtained. Table 3 shows all the experimental area covered by these 3 designs and their associate values for the true response and for the fitted one.

TABLE 3: The layout and main results of the three sequential designs.

Design Point	x_1	x_2	$\pi(\mathbf{x}, \boldsymbol{\beta})$	M	y	pr
C.1	1.20323	−1.40966	0.04581	100	5	0.05
D.1	1.40966	−1.40966	0.01360	100	2	0.02
E.1	1.40966	−1.20323	0.04581	100	6	0.06
F.1	1.20323	−1.20323	0.13312	100	13	0.13
O.1	1.30645	−1.30645	0.04675	100	7	0.07
C.2	1.17791	−1.13558	0.19745	100	11	0.11
D.2	1.38435	−1.13558	0.07473	100	6	0.06
E.2	1.38435	−0.92915	0.18168	100	19	0.19
F.2	1.17791	−0.92915	0.38311	100	38	0.38
O.2	1.28113	−1.03237	0.19273	100	27	0.27
C.3	1.14982	−0.86177	0.47800	100	49	0.49
D.3	1.35626	−0.86177	0.25623	100	34	0.34
E.3	1.35626	−0.65534	0.42746	100	44	0.44
F.3	1.14982	−0.65534	0.64570	100	66	0.66
O.3	1.25304	−0.75856	0.45789	100	46	0.46

It can be seen on Table 3 that the sequential generation of successes increase from the first design to the last one (see column labeled as y), just as the migration from the first experimental region to the third one. This situation is coherent in the sense of the sequential exploration of the true response and the sequential designs that the experimenter conducts to successive "best" approximations to the maximum. Using the glm() function of R, the fitted model for all these 15 design points is shown on Table 4. The fitted linear predictor for this model can be expressed as $\widehat{\eta} = 9.750339 - 4.59569 x_1 + 5.137089 x_2$.

TABLE 4: The ANOVA table for the fitted model of the 15 design points.

	Estimate	Std. Error	z value	$Pr(> |z|)$
(Intercept)	9.750339	0.6450672	15.115229	0.000
x.01	−4.595690	0.4332252	−10.608014	0.000
x.02	5.137089	0.6567979	7.821416	0.000
x.12	0.000000	0.0000000	0.000000	0.000
x.11	0.000000	0.0000000	0.000000	0.000
x.22	0.000000	0.0000000	0.000000	0.000

4.2.2. Several Experimental Runs

So far, we have defined a *strategy of experimentation* consisting of the definition of 3 successive two level-factorial designs for single levels of w, L and S, respectively. Linking all the 3 designs in the way described before, the whole planned budget is used at the end of the third fitted model. Thus, a complete second order

linear predictor for the logistic model is obtained for all the 15 design points. As stated before, this set of 15 design points constitutes one "case study" of design points.

In order to consider some variation on the observed response, now we run 15 times each case study using the same levels of w, L and S. That is, for every sequence of 5 points-designs, a random number of successes (and then, proportions) is generated, so that different paths of approximation to maxima conditions are obtained. For each one of these paths, a scheme of 3 sequential designs is also obtained, and then, a respective second order complete model is fitted. Therefore, if a design strategy is defined by a single row of the $\mathbf{LS}(0.05)$ matrix, its i-th row works as the generator of 15 random approximation paths to the maxima conditions. Associated to each one, a fitted model of maximum $p = 6$ parameters is also obtained. For simplicity, we call every single experimental run as "one case study", whereas all the 15 experimental runs are labeled as "fifteen case studies" or "one data frame", as well. The next natural step is to compute the two criteria of evaluation, T_1 and T_2, as they were defined in expressions (11) and (12), respectively, for a single data frame.

1. *First criterion.* We have that the observed FIM for a single data frame is given by $\sum_{i=1}^{15} \log \det_i(\mathbf{X}'\mathbf{W}\mathbf{X})$. The notation $\det_i(\mathbf{X}'\mathbf{W}\mathbf{X})$ states that for each one of the 15 cases study considered, there are particular values for the design matrix, $\mathbf{X}$, and also for the matrix of "weights", $\mathbf{W}$. So, there is also an associated FIM for every one of the 15 fitted models, whose determinant is the one described before. For each one of the 15 cases study, the estimated matrix of "weights", $\widehat{\mathbf{W}}_i$, has the form $\widehat{\mathbf{W}}_i = \mathrm{diag}\{m_i\widehat{\pi}_i(1-\widehat{\pi}_i)\}$, where $\widehat{\pi}_i = \widehat{\pi}(\mathbf{x}_i, \widehat{\boldsymbol{\beta}})$ is calculated on the last model of each case study. We determine $\widehat{\pi}$ by the consideration of its maximum likelihood estimation, which is the proportion of observed successes. That is: $\widehat{\pi}_i = \frac{y_i}{100}$. Therefore, for every case study, the observed FIM is calculated following this way. We can extend this procedure to all the "data frame" by means of the calculation of the same measure for every "case study" and then calculating the average of them, that is $T_1 = \sum_{i=1}^{15} \det_i(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})$. Table 5 shows the calculation of T_1 for all the 15 case studies generated from w_1, L_1 and S_1, that can be seen on the second column of the same table.

2. *Second criterion.* For this second criterion, the calculation of T_2 is completely analogous as in the first one. For each one of the 15 case studies, the quantity $\pi(\widehat{\mathbf{x}}, \boldsymbol{\beta})$ is shown on the third column of Table 5. On the determination of each one of them, the following steps has been made: (a) the vector of estimates of the unknown parameters, $\widehat{\boldsymbol{\beta}}$; (b) the value of $\mathbf{x}$ that makes the fitted model $\widehat{\pi}(\mathbf{x}, \widehat{\boldsymbol{\beta}})$ to reach a maximum, which is $\widehat{\mathbf{x}}$; and (c) the value of the real process evaluated on $\widehat{\mathbf{x}}$, that is $\pi(\widehat{\mathbf{x}}, \boldsymbol{\beta})$.

As we can see in Table 5, this design strategy can be summarized on both criteria: $T_1 = 0.818991$ and $T_2 = 0.706840$. These indicators condense the amount of information that can be obtained by choosing levels w_1, L_1 and S_1 for the three

TABLE 5: The calculation of T_1 and T_2 for all the 15 cases study generated by $w = w_1$, $S = S_1$ and $L = L_1$.

Case Study	T_1	T_2
SC.01	0.664814	0.898778
SC.02	1.481098	0.682047
SC.03	0.849250	0.628167
SC.04	0.856407	0.899190
SC.05	0.680511	0.445337
SC.06	0.961489	0.756512
SC.07	0.584377	0.927986
SC.08	0.587970	0.748809
SC.09	-0.066011	0.658838
SC.10	1.937709	0.652120
SC.11	0.678024	0.714656
SC.12	0.677629	0.611532
SC.13	0.619965	0.628572
SC.14	1.041491	0.579345
SC.15	0.729993	0.770705
Mean	0.818991	0.706840

evaluation variables on the attainment of optima conditions for the process given by Equation (13). These measurements can give an insight on how good the design strategy selected, has been.

By observing both values of T_1 and T_2, they have no information itself in the sense that they are measurements of only one design strategy, that is the one given by w_1, L_1 and S_1. Now, the next natural step may suggest to consider all the rest of rows of $\mathbf{LS}(0.05)$ matrix and to calculate both T_1 and T_2 for each one. This may lead us to have a more precise idea of how this design strategy –the one that takes $\mathbf{LS}(0.05)$– works.

4.3. An Example Considering a Complete LS Matrix

This section presents a generalization of what has been done before using one row of the $\mathbf{LS}(w)$ matrix. That is, the determination for both T_1 and T_2 now is performed the same way as in the last section. We now perform 15 experimental runs for every one of the rows of $\mathbf{LS}(0.05)$ matrix, then calculating both values T_1 and T_2 for each one of the 25 data frames contained on the $\mathbf{LS}(0.05)$ matrix.

For a clearer representation of how T_1 and T_2 behave on all 25 rows of $\mathbf{LS}(0.05)$, we build a 2-dimensional grid chart, which consists of two orthogonal axes, named L and S, that are divided into 5 levels the same way as shown on (14) and (15). Focused on the 25-points grid, in which every one is a pair of the form $L_i S_i$, for $i = 1, \ldots, 5$, then the idea is to build two charts, one for each T_1 and T_2, so that the value of both quantities are shown on them. For example, every row of the $\mathbf{LS}(0.05)$ matrix corresponds to one design strategy, that is summarized on both T_1 and T_2. Thus, if all the 5×5 points –that are the same 25 rows of the matrix– are represented on the corresponding 5×5 grid, then the value of T_1 for each row is shown each point, and similarly to T_2. Therefore, and for a single matrix of the

form $\mathbf{LS}(w)$, one grid chart made this way is a useful tool to compare all the values of T_1 (and also T_2), so that it can be seen there what are the coordinates on the chart that leads to the maximum value of it. In summary, it is analogous to find the design strategy –given by a certain pair L_iS_i– that makes the maximum value of the amount of information obtained from each one of the 25 design strategies considered on a single $\mathbf{LS}(w)$ matrix. For example, when $w = 0.05$, both grid charts of $T_1(L, S)$ and $T_2(L, S)$ are shown on Figure 5.
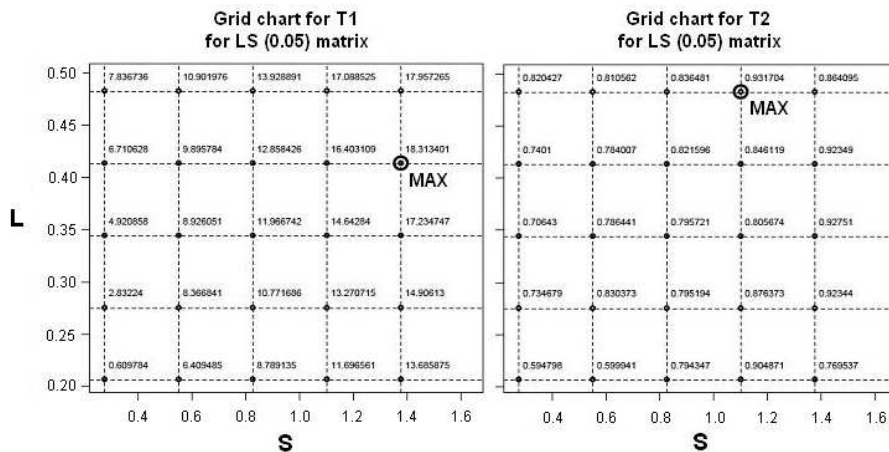

FIGURE 7: A representation of grid charts for $\mathbf{LS}(0.05)$. In both charts, the values for L and S that correspond to maxima values for T_1 and T_2 are labeled as "MAX".

Figure 7 shows that higher values for T_1 and T_2 are obtained for higher levels for both L and S when $w = 5\%$. Since each one has been calculated for 15 runs, we finally find that the maximum value for the first criterion is $T_1 = 18.313401$, for $L = L_4$ and for $S = S_5$. For the second criterion, we have $T_2 = 0.931704$, for $L = L_5$ and $S = S_4$.

4.4. An Extended Example Considering Five LS Matrices

After evaluating a complete $\mathbf{LS}$ matrix for $w = 0.05$, and since we have defined 5 levels for w, we have calculated all the 5×5 pairs L_iS_i for the five corresponding $\mathbf{LS}$ matrices, that is, for $\mathbf{w} = (0.05, 0.10, 0.15, 0.20, 0.25)'$. As a result, and for every one of these 5 matrices, a grid-chart similar to Figure 7 is obtained. In each chart, the maximum value of both T_1 and T_2 is also determined. We can summarize now all these in two new charts, one for each T_1 and T_2, on which we present the maxima values obtained for both measurements, for every one of the levels considered for w.

Figure 8 shows maxima values for both T_1 and T_2, evaluated on all the levels of w considered. On this figure, it can be seen that higher values of $\max(T_1)$ and $\max(T_2)$ are obtained for small levels of w. In particular, when $w = 0.10$ values of both measurements rises their maxima values, that are $\max(T_1) = 19.37509$ and
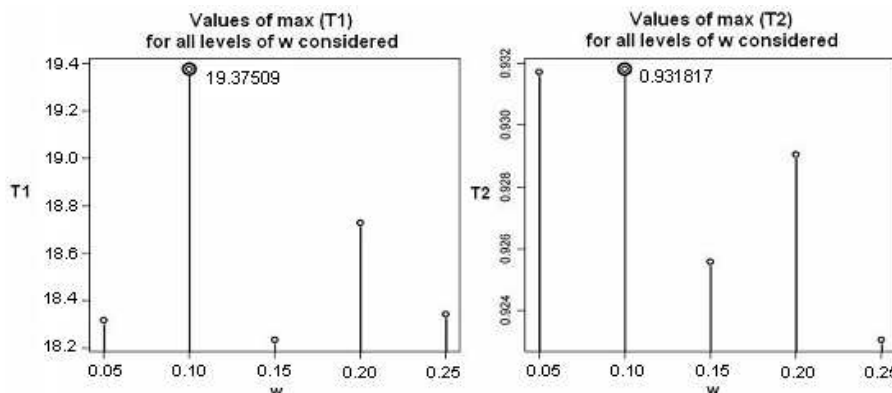
FIGURE 8: A comparison among values of $\max(T_1)$ and $\max(T_2)$ for $\mathbf{LS}(w)$ matrices, when $\mathbf{w} = (0.05, 0.10, 0.15, 0.20, 0.25)'$.

$\max(T_2) = 0.931817$. An important fact is that both criteria coincide on attaining a maximum value for smaller levels of w. This situation seems to be more clear on the left side of Figure 8.
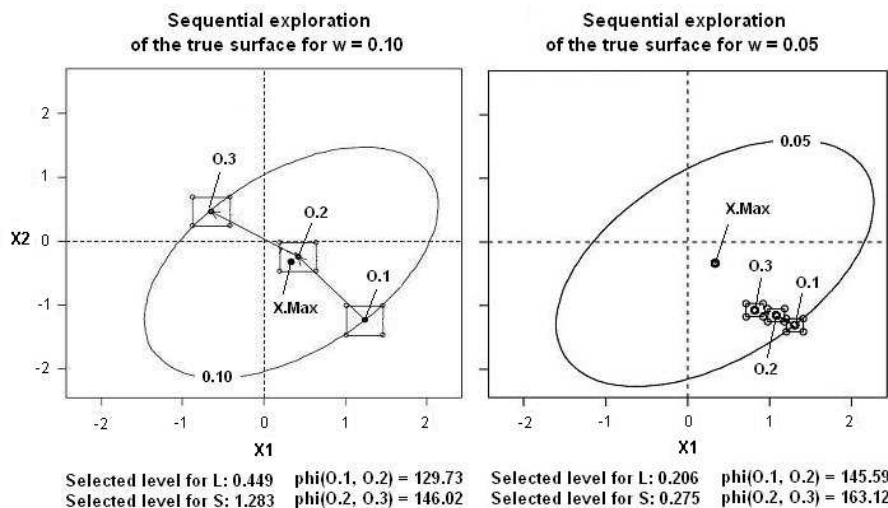

FIGURE 9: Two typical sequential designs resulting for two design strategies: one with $w = 5\%$ (on the left) and the other one with $w = 10\%$ (on the right).

4.4.1. Comparison of Two Cases

Provided that "best" values for both measurements are obtained for small levels of w, we now present two figures of how the exploration of the true process evolved from the first factorial design to the third one. When L and S take their maxima

levels among all considered, that is $L = 0.449$ and $S = 1.283$, and when $w = 0.10$, the resultant experimental path is presented on Figure 9. On the left side of this figure, it appears as the aspect of the row of the corresponding $\mathbf{LS}(0.10)$ matrix defined by $L = L_5$ and $S = S_5$, that also lead to the highest value for both T_1 and T_2. On the right side of the same figure we present the evolution of the design strategy for $w = 0.05$, when L and S take the levels $L = 0.206$ and $S = 0.275$, respectively.

5. Conclusions

By looking at both charts of Figure 9, some interesting observations can be made. From the point of view of the selection of levels of w, the case of $w = 0.10$ appears to be "better" than the one on that $w = 0.05$ in the sense that the first one has a wider range of area of the true response covered by the three designs. This situation may suggest that the amount of information on the surface is more spread than in the case where $w = 0.05$, whose experimental area is smaller. On the left side of the same figure, the fact that the third factorial designs "passes" through the maximum of the true surface —labeled as X.Max— does not seem to be a problem since the main goal of the exploration is not to fall into the neighborhoods of the maximum but to gain the best information about the design. In this sense, the right side of the same figure shows a more congregated display for experimental points, while the left one has a wider spread of them.

The situation for the "best" values obtained can be summarized by $\boldsymbol{\psi}^* = (w, L, S)' = (0.05, 0.449, 1.283)'$. As we said before, this value of $\boldsymbol{\psi}$ also gives the best configuration for design strategies among all the ones evaluated so far that were performed on the evaluation of the five matrices of the form $\mathbf{LS}(w)$. We can see that among all possible values T_1 and T_2 can take within the levels of w, L and S, a common configuration of $\boldsymbol{\psi}^*$ seems to give the value in which both measurements T_1 and T_2 reach their maxima values. This value of ψ leads to define $D^*(\boldsymbol{\psi})$, which is the best design strategy found for the example in exam. A classical RSM–like experimentation was followed successfully by sequentially exploring the process, modeled by the GLM scope.

Finally, we summarize all that we have said before in the following way: *"relating a RSM problem with the definition of three evaluation variables, given by vector $\boldsymbol{\psi} = (w, L, S)'$, a sequential design strategy for logistic models, $D(\boldsymbol{\psi})$, reflects the amount of information obtained for the true process when a certain experimental budget has been used. The measure of this amount of information is condensed by the definition of two criteria, T_1 and T_2 for an adjusted model that contains all the information of the true surface. Therefore, and since T_1 and T_2 depend on the values of $\boldsymbol{\psi}$, the goal of the experiment is to compare different levels of $\boldsymbol{\psi}$ in order to obtain those which $T_1[D(\boldsymbol{\psi})] \to \max$ and $T_2[D(\boldsymbol{\psi})] \to \max$. Values of $\boldsymbol{\psi}$ that satisfy this condition are said to lead the experiment to the 'best' design strategy of exploration for the true process. As a result, highest levels of L and S and a common level of w satisfy both criteria simultaneously".*

Acknowledgements

References

Atkinson, A. C. (2006), *Generalized Linear Models and Response Transformation*, Khuri, chapter 8.

Box, G. E. P. & Draper, N. R. (1987), *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York, United States.

Box, G. E. P. & Draper, N. R. (2007), *Response Surfaces, Mixtures and Ridge Analysis*, second edn, John Wiley & Sons, New York, United States.

Box, G. E. P., Hunter, W. G. & Hunter, J. S. (2005), *Statistics for Experimenters: Design, Innovation and Analysis*, second edn, John Wiley & Sons, New York, United States.

Box, G. E. P. & Wilson, K. B. (1951), 'On the Experimental Attainment of Optimum Conditions', *Journal of the Royal Statistical Society* **Series B**(13), 1–45.

Collett, D. (2002), *Modelling Binary Data*, second edn, Chapman & Hall/CRC, Boca Raton, United States.

Cordeiro, G. M. & De Andrade Lima Neto, E. (2004), Modelos paramétricos, *in* '16o. SINAPE', Recife, Brasil.

Cox, D. R. & Reid, N. (2000), *The Theory of the Design of Experiments*, Chapman & Hall/CRC, Boca Raton, United States.

De Zan, A. T. (2006), Principios de metodología de superficie de respuesta para modelos logísticos, Ph.D. Thesis, Departament d' Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, España.

Dobson, A. J. (2002), *An introduction to Generalized Linear Models*, second edn, Chapman & Hall/CRC, Boca Raton, United States.

Firth, D. (1991), *Generalized Linear Models*, Hinkley, et al., chapter 3.

Khuri, A. I. (1993), 'Response Surface Methodology within the Framework of GLM', *Journal of Combinatorics Information & System Sciences* **18**, 193–202.

Khuri, A. I. (2001), 'An Overview of the Use of Generalized Linear Models in Response Surface Methodology', *Nonlinear Analysis* **47**, 2023–2034.

Khuri, A. I. (2006), *Response Surface Methodology and Related Topics*, World Scientific Publishing, New Jersey, United States.

Khuri, A. I. & Cornell, J. A. (1996), *Response Surfaces: Designs and Analyses*, second edn, M. Dekker, New York, United States.

Khuri, A. I. & Mukhopadhyay, S. (2006), *GLM Designs: The Dependence on Unknown Parameters Dilemma*, Khuri, chapter 9.

Lewis, S. L., Montgomery, D. C. & Myers, R. H. (2001*a*), 'Confidence Interval Coverage for Designed Experiments with GLMs', *Journal of Quality Technology* **33**, 279–292.

Lewis, S. L., Montgomery, D. C. & Myers, R. H. (2001*b*), 'Examples of Designed Experiments with Nonnormal Responses', *Journal of Quality Technology* **33**, 265–278.

Lindsey, J. K. (1997), *Applying Generalized Linear Models*, Springer, New York, United States.

McCullagh, P. M. & Nelder, J. A. (1989), *Generalized Linear Models*, second edn, Chapman & Hall/CRC, Boca Raton, United States.

Montgomery, D. C. (2005), *Design and Analysis of Experiments*, sixth edn, John Wiley & Sons, New York, United States.

Myers, R. H. (1999), 'Response Surface Methodology –Current Status and Future Directions', *Journal of Quality Technology* **31**, 30–44.

Myers, R. H. & Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, second edn, John Wiley & Sons, New York, United States.

Myers, R. H., Montgomery, D. C. & Vining, G. G. (2001*a*), *Generalized Linear Models: With Applications in Engineering and the Sciences*, John Wiley & Sons, New York. United States.

Myers, R. H., Montgomery, D. C. & Vining, G. G. (2001*b*), *Regression Analysis. Theory, Methods and Applications*, Springer, New York, United States.

Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M. & Kowalski, S. M. (2004), 'Response Surface Methodology: A Retrospective and Literature Survey', *Journal of Quality Technology* **36**, 53–77.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized Linear Models', *Journal of the Royal Statistical Society* **Series A**(135), 370–384.

Pukelsheim, F. (1993), *Optimal Design of Experiments*, John Wiley & Sons, New York, United States.

R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Robinson, K. S. & Khuri, A. I. (2003), 'Quantile Dispersion Graphs for Evaluating and Comparing Designs for Logistic Regression Models', *Computational Statistics & Data Analysis* **43**, 47–62.