

New methods for seeking optimum conditions

Nuovi metodi per la ricerca delle condizioni di ottimalità

Valeria Sambucini

Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università di Roma “La Sapienza”
valeria.sambucini@uniroma1.it

Riassunto: Si propongono nuove procedure inferenziali per la ricerca del punto di ottimo di una superficie di risposta. In tali procedure le coordinate del punto stazionario della funzione di risposta vera vengono introdotte come parametri di interesse. Ciò rende possibile, in alternativa ai metodi correnti, sia l'uso della verosimiglianza profilo, sia il ricorso a procedure Bayesiane per l'esplorazione sequenziale della superficie di risposta.

Keywords: Bayesian analysis, central composite designs, HPD regions, profile likelihood, reference priors, response surface methodology, rotatability.

1. Introduction

Let us consider a response variable Y related to k continuous factors $U_1, \dots, U_k$. Response surface methodology (RSM), introduced by Box and Wilson (1951), aims at finding the combination of factor levels which produces the optimal response. Detailed references for these techniques include Box and Draper (1987) and Myers and Montgomery (1995).

For computational convenience, the original input variables are usually converted to coded variables, $X_1, \dots, X_k$, standardized so that the design center is at the point $(x_1, \dots, x_k) = \mathbf{0}$. Moreover the unknown relation between the response and the coded factors, i.e. true response function, is approximated by a low-order polynomial model.

RSM is generally implemented in two phases. The main purpose of the first phase is to determine whether the current levels of the input factors are close to the optimum point or far from it. The standard procedure for seeking the optimum region is the method of steepest ascent (or descent). Suppose that we are searching for the maximum point of the response surface. A first order model is fitted, generally using a 2^k factorial design. Runs at the center point of the experimental region are added so that a curvature check of the underlying surface can be made. If there is no indication of curvature, several runs are conducted along the path of steepest ascent until no additional increase in response is evident. The location on the path where a maximum response is observed will be taken as the center point of the next experiment. If curvature is found to be quite evident, a more elaborate experiment should be conducted and the second phase of RSM begins. Often, the first order plan is expanded to a second order design by adding $2k$ axial points to make it a central composite design (CCD). This class of designs was proposed by Box and Wilson (1951) as a natural alternative to the 3^k factorial design for its requirement of fewer experimental runs and its flexibility. Many response surface designs are considered in the statistical literature as alternatives to CCD's: between the most recent proposals see, for instance, Mee (2001) and Trinca and Gilmour (2001).

The main goal of the second phase of RSM is a detailed analysis of the second order response function. The second order polynomial model can be compactly written as

$$M_P: \quad y = \beta_0 + \mathbf{x}^T \boldsymbol{\beta} + \mathbf{x}^T \mathbf{B} \mathbf{x} + \varepsilon, \quad (1)$$

where ε is the random error which is assumed to be distributed as $N(0, \sigma^2)$, with σ^2 unknown. Here $\mathbf{x}$ is a $k \times 1$ vector of factor levels, β_0 is the intercept term, $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients β_i and $\mathbf{B}$ is a $k \times k$ symmetric matrix of regression coefficients with i -th diagonal element equal to β_{ii} and the (ij) -th off-diagonal element equal to $(1/2)\beta_{ij}$. First the estimated stationary point, $\hat{\mathbf{x}}^S = -\frac{1}{2}\hat{\mathbf{B}}^{-1}\hat{\boldsymbol{\beta}}$, is computed, where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{B}}$ are the maximum likelihood estimates (MLE) of $\boldsymbol{\beta}$ and $\mathbf{B}$. Then graphical and analytic procedures are used to determine the nature of $\hat{\mathbf{x}}^S$ and to characterize the behavior of the response surface in the experimental region $\mathcal{R}$. Furthermore, if $\hat{\mathbf{x}}^S$ is a point of maximum and lies inside $\mathcal{R}$, confidence regions on the location of the true stationary point are derived (see Box and Hunter (1954)). On the contrary, if $\hat{\mathbf{x}}^S$ is a saddle point or a point of maximum that resides outside $\mathcal{R}$, a further exploration is suggested.

The new proposed parametrization of model (1) is

$$M_R: \quad y = \alpha_0 + (\mathbf{x} - \boldsymbol{\xi})^T \mathbf{A} (\mathbf{x} - \boldsymbol{\xi}) + \varepsilon, \quad (2)$$

where $\alpha_0 = \beta_0 - \frac{1}{4}\boldsymbol{\beta}^T \mathbf{B}^{-1} \boldsymbol{\beta}$, $\boldsymbol{\xi} = -\frac{1}{2}\mathbf{B}^{-1} \boldsymbol{\beta}$ and $\mathbf{A} = \mathbf{B}$. Since these relations are invertible, models (1) and (2) are equivalent and the MLE's of the parameters of M_R can be derived, through invariance, from the MLE's of the parameters of M_P . Remark that $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)$ represents the stationary point of the true response function, α_0 is the value of the true response function on $\boldsymbol{\xi}$ and $\mathbf{A}$ is a $k \times k$ symmetric matrix of coefficients with diagonal elements α_i and off-diagonal elements $(1/2)\alpha_{ij}$. Therefore we can introduce likelihood techniques (see Section 2) and Bayesian procedures (see Section 3) to make inference on the parameter of interest $\boldsymbol{\xi}$, treating the other parameters of model M_R as nuisance parameters.

2. The profile likelihood for $\boldsymbol{\xi}$

Let us suppose that n experimental runs are taken on various combination of x -s. The model M_R can be conveniently written in the form

$$M_R: \quad \mathbf{y} = \mathbf{X}_{\boldsymbol{\xi}} \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (3)$$

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_k, \alpha_{12}, \dots, \alpha_{k-1,k})$ and

$$\mathbf{X}_{\boldsymbol{\xi}} = \begin{bmatrix} 1 & (x_{11}-\xi_1)^2 & \dots & (x_{1k}-\xi_k)^2 & (x_{11}-\xi_1)(x_{12}-\xi_2) & \dots & (x_{1,k-1}-\xi_{k-1})(x_{1k}-\xi_k) \\ 1 & (x_{21}-\xi_1)^2 & \dots & (x_{2k}-\xi_k)^2 & (x_{21}-\xi_1)(x_{22}-\xi_2) & \dots & (x_{2,k-1}-\xi_{k-1})(x_{2k}-\xi_k) \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & (x_{n1}-\xi_1)^2 & \dots & (x_{nk}-\xi_k)^2 & (x_{n1}-\xi_1)(x_{n2}-\xi_2) & \dots & (x_{n,k-1}-\xi_{k-1})(x_{nk}-\xi_k) \end{bmatrix}.$$

It is easily proved that the normed profile likelihood for the parameter of interest is

$$\bar{L}_{\max}(\boldsymbol{\xi} | \mathbf{D}) = \frac{L_{\max}(\boldsymbol{\xi} | \mathbf{D})}{L_{\max}(\hat{\boldsymbol{\xi}} | \mathbf{D})} = \left[\frac{(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}} \hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}))^T (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}} \hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}))}{(\mathbf{y} - \mathbf{X}_{\hat{\boldsymbol{\xi}}} \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\xi}}))^T (\mathbf{y} - \mathbf{X}_{\hat{\boldsymbol{\xi}}} \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\xi}}))} \right]^{-\frac{n}{2}}, \quad (4)$$

where $\mathbf{D}$ is the data matrix and $\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = (\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}})^{-1} \mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{y}$. Assuming that $\boldsymbol{\xi}$ is the true maximum point, we will consider (4) only for the values of $\boldsymbol{\xi}$ such that the Hessian matrix of the estimated response surface written as function of $\boldsymbol{\xi}$, i.e. $2\hat{\mathbf{A}}(\boldsymbol{\xi})$, is negative definite.

In both phases of RSM, when the observed data suggest an additional exploration, we propose to use the information provided by the q -level profile likelihood sets to locate the next experimental region. Note that the method of steepest ascent (in the first phase) does not determine the distance from the center of the current experiment at which to conduct the next experiment. Hence additional experimental runs along the path are arbitrarily chosen by the researcher. On the contrary the likelihood method allows the researcher to locate directly the new experimental region and provides information about the range to be fixed for each factor.

3. Bayesian analysis of the model M_R

In order to have a non-informative prior, we resort to the well-known reference priors method (see, for instance, Berger and Bernardo (1992)). The conditional reference prior for the nuisance parameter (α, σ^2) given the parameter of interest ξ turns out to be $\pi^R(\alpha, \sigma^2 | \xi) \propto 1/\sigma^2$. This result does not depend on the order in which the parametrisation is taken. The marginal reference prior for ξ is given by

$$\pi^R(\xi) \propto |\mathbf{X}_\xi^T \mathbf{X}_\xi|^{-1/2}. \quad (5)$$

The reference prior (5) is improper and depends on the experimental design. Note that, using second order model, a popular property to take into account is rotatability. A design is said to be rotatable if the variance of the estimated response, $\hat{y}(\mathbf{x})$, depends only on the distance of $\mathbf{x}$ from the design center and not on the direction. Therefore, a rotatable design guarantees that all the points in the factor space at the same distance from the origin are treated as being “equally important”. When $k = 2$, it is possible to show that if we use a rotatable CCD the corresponding distribution $\pi^R(\xi)$ is constant on circles around the design center. Thus these points are considered *a priori* as “equally probable”. This result does not hold if we adopt a 3^k factorial plan, which is orthogonal, but not rotatable. However, in both cases, $\pi^R(\xi)$ presents its unique mode at the design center, $\xi = (0, 0)$.

If prior information about the location of the true stationary point is available, it is useful to introduce an informative prior distribution for ξ . Always assuming the conditional reference prior $\pi^R(\alpha, \sigma^2 | \xi) \propto 1/\sigma^2$ for the nuisance parameters, some possible choice for the informative marginal prior of ξ are the normal, the generalized Student-T and the uniform distribution on the experimental region, that is:

$$(i) \pi^I(\xi) = N_k(\mu_\xi, \Sigma_\xi), \quad (ii) \pi^I(\xi) = T_k(\mu_\xi, \Sigma_\xi; \nu), \quad (iii) \pi^I(\xi) \propto 1, \text{ for } \xi \in \mathcal{R}.$$

The specification of the prior hyperparameters should be coherent with the properties of the employed experimental design. For example, in both cases (i) and (ii), when a rotatable design has been chosen, it is reasonable to assume $\mu_\xi = \mathbf{0}$ and Σ_ξ diagonal, so that the rotatability is ensured from a Bayesian point of view as well. The diagonal elements of Σ_ξ can be fixed with reference to the prior probability assigned to $\mathcal{R}$.

Whether we assume for ξ a non-informative prior or an informative one, integrating out the nuisance parameters from the joint posterior distribution of (ξ, α, σ^2) , we find the posterior marginal distribution of the true stationary point

$$\pi(\xi | \mathbf{D}) \propto \pi(\xi) \left[(\mathbf{y} - \mathbf{X}_\xi \hat{\alpha}(\xi))^T (\mathbf{y} - \mathbf{X}_\xi \hat{\alpha}(\xi)) \right]^{-\frac{n-p'}{2}} \quad (6)$$

where $p' = 1 + k + \frac{k(k-1)}{2}$ is the dimension of the vector $\hat{\alpha}(\xi)$. As before, when ξ is assumed to be a maximum point, we will consider the density (6) only for the values of ξ such that the matrix $2\hat{A}(\xi)$ is negative definite.

Some graphical examples with $k = 2$ and $k = 3$, not reported here, through MCMC simulations, show the behavior of the 90 and 95% HPD regions using informative and non-informative priors. These HPD regions are compared with the standard frequentist regions for the true stationary point, in terms of shapes and widths (see Sambucini (2003)). Further simulations evaluate the underlying coverage probabilities of the 95% HPD regions, when $k = 2$. Fixed α , three coded true response functions were considered, with maximum points, respectively, at $(0, 0)$, $(0.3, 0.3)$ and $(0.7, 0.7)$. For each true response function, samples of size 12 were simulated from a rotatable CCD with 4 center points, using $\sigma^2 = 1$, until 1000 data sets, such that $\hat{\xi}$ is a maximum point, were obtained. The coverage rates of the 95% HPD regions, shown in Table 1, were computed, through MCMC methods, using three different prior distributions: the reference prior, an informative normal distribution which assigns a prior probability to $\mathcal{R}$ approximately equal to 0.95 ($\pi_1^I(\xi)$) and the informative uniform distribution on $\mathcal{R}$ ($\pi_2^I(\xi)$).

Table 1: Coverage rates of 95% HPD regions, when $\sigma^2 = 1$ and $\alpha = (100, -2.5, -5, 3.75)$.

	$\pi^R(\xi)$	$\pi_1^I(\xi) = N_2(\mathbf{0}, \text{diag}(0.33, 0.33))$	$\pi_2^I(\xi) \propto 1, \xi \in \mathcal{R}$
$\xi = (0, 0)$	0.962	0.965	0.963
$\xi = (0.3, 0.3)$	0.968	0.970	0.961
$\xi = (0.7, 0.7)$	0.955	0.930	0.940

In general, the observed coverage probabilities are reasonably close to the nominal value, showing the good frequentist behavior of the Bayesian procedure. However, when the ξ is assumed to be near the boundary of $\mathcal{R}$ and we use the informative priors, the coverage rates are slightly lower than the nominal level. This occurs, in particular, when we assume for ξ the normal informative prior ($\pi_1^I(\xi)$), which is concentrated around the center design.

References

- Berger J.O. and Bernardo J.M. (1992) On the development of reference priors (with discussion), *In Bayesian Statistics 4*, Ed. J.M. Bernardo et al., Oxford Univ. Press.
- Box G.E.P. and Draper N.R. (1987) *Empirical Model Building and Response Surfaces*, Wiley, New York.
- Box G.E.P. and Hunter J.S. (1954) A confidence region for the solution of a set of simultaneous equations with an application to experimental design, *Biometrika*, 41, 190–199.
- Box G.E.P. and Wilson K.B. (1951) On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society, Series B*, 13, 1–45.
- Mee R. (2001) Non central composite designs, *Technometrics*, 43, 34–43.
- Myers R.H. and Montgomery D.C. (1995) *Response Surface Methodology: Process and Produce Optimization Using Designed Experiments*, Wiley, New York.
- Sambucini V. (2003) *Nuovi metodi per l'esplorazione delle superfici di risposta*, Tesi di Dottorato, Dip. di Statistica, Probab. e Stat. Appl., Univ. di Roma, “La Sapienza”.
- Trinca L.A. and Gilmour S.G. (2001) Multistratum response surface designs, *Technometrics*, 43, 25–33.