Likelihood and Bayesian techniques for the analysis of a response surface

Valeria Sambucini and Ludovico Piccinato
Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università "La Sapienza", Piazzale Aldo Moro 5, Roma 00185, Italia

Abstract

In response surface methodology a second order polynomial model is typically used to make inferences on the stationary point $\boldsymbol{\xi}$ of the true response function. The standard confidence regions for the true stationary point are due to Box and Hunter (1954). We introduce an alternative parametrization, in which $\boldsymbol{\xi}$ appears as parameter of interest. This way likelihood techniques and Bayesian analysis are more easily performed. An empirical method to get HPD regions for the true maximum point (not simply for the stationary point) is also proposed and a simulation study is produced to compare their coverage probabilities and sizes with the coverage probabilities and sizes of the frequentist regions.

Keywords: Bayesian analysis, confidence regions, HPD regions, integrated likelihood, MCMC simulations, profile likelihood, response surface methodology, rotatability.

1 Introduction

Response Surface Methodology (RSM), first introduced by Box and Wilson (1951), is typically used to find the levels of k continuous factors which optimize a response variable, Y. This set of techniques is described in detail in many textbooks, including those by Davies (1960), Box and Draper (1987), Khuri and Cornell (1987) and Myers and Montgomery (1995). The relevance for biometric and pharmaceutical research was recently stressed by Peterson, Cahya and del Castillo (2002).

The original input variables are usually converted to coded variables, $X_1, ..., X_k$, so that the design center is at the point $(x_1, ..., x_k) = \mathbf{0}$. Moreover the true response function, $\varphi(x_1, x_2, ..., x_k)$, i.e. the unknown relationship between the response variable and the explanatory variables, is usually approximated by a second order polynomial model. This approximation can be considered reliable in the experimental region, $\mathcal{R}$, that is the sub-region of the factor space over which the experiment is performed.

The second order polynomial model can be compactly written as

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

where ε is the random error which is assumed normally distributed with zero mean and unknown variance σ^2. 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 $(1/2)\beta_{ij}$. Model

(1) is characterized by $p+1$ unknown parameters, where $p = 1 + 2k + \frac{k(k-1)}{2}$ is the number of regression coefficients, including the intercept term.

The exploration of the fitted second order surface, $\hat{y}$, is usually based on the estimation of the true stationary point, $\hat{\mathbf{x}}^S = -\frac{1}{2}\hat{\mathbf{B}}^{-1}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{B}}$ are the maximum likelihood estimates (MLEs) of $\boldsymbol{\beta}$ and $\mathbf{B}$. Depending on the sign of the eigenvalues of $\hat{\mathbf{B}}$, $\hat{\mathbf{x}}^S$ could be a maximum, a minimum or a saddle point of $\hat{y}$. On the other hand, the magnitude of the eigenvalues characterize the degree of curvature of the fitted quadratic response: an eigenvalue near 0 indicates the presence of a ridge in the surface. Therefore, a *canonical analysis* can be performed to determine the nature of $\hat{\mathbf{x}}^S$ and to interpret the behavior of the response system in the experimental region. Furthermore, in order to make inferences on the nature of the response surface, Carter *et al* (1986 and 1990), Peterson (1993) and Bisgaard and Ankenman (1996) give methods to get standard errors and large-sample confidence intervals for the eigenvalues of $\mathbf{B}$.

Let us suppose that we are interested in the point of maximum response. If $\hat{\mathbf{x}}^S$ is not a maximum point or if it is a maximum that lies outside the experimental region, it would not be reasonable to suggest it as a candidate for optimum conditions, but attention should be focused on the nature of the system inside $\mathcal{R}$, where the fitted model is reliable, and further experiments should be carried out. In these situations it is useful to employ a *ridge analysis* (see, for instance, Draper, 1963), which locates the points of maximum response at a fixed distance from the design center. This methodology provides information about the role of the design variables inside $\mathcal{R}$ and the areas where future experiments could be made.

Confidence regions are obviously useful to locate at least approximately the true maximum point. It must be stressed that they are especially interesting and reliable when $\hat{\mathbf{x}}^S$ is a maximum for the fitted surface and resides inside $\mathcal{R}$. Box and Hunter (1954) derive a confidence region for the true stationary point. Denoting with $\boldsymbol{\xi} = (\xi_1, ..., \xi_k)$ the unknown stationary point and with $\hat{\boldsymbol{d}}(\mathbf{x}) = \hat{\boldsymbol{\beta}} + 2\hat{\mathbf{B}}^T\mathbf{x}$ the vector of derivatives $\partial\hat{y}(\mathbf{x})/\partial x_j$, for all $j = 1, ..., k$, the $100(1-\alpha)\%$ Box-Hunter confidence region for $\boldsymbol{\xi}$ is given by

$$C_{BH} = \left\{\boldsymbol{\xi} : \boldsymbol{d}(\boldsymbol{\xi})^T\, \mathbf{V}^{-1}_{\boldsymbol{d}(\boldsymbol{\xi})}\, \boldsymbol{d}(\boldsymbol{\xi}) \leq k\, F(k, n-p\,; \alpha)\right\}, \tag{2}$$

where $\mathbf{V}_{\boldsymbol{d}(\boldsymbol{\xi})}$ is the estimate of the variance-covariance matrix of $\boldsymbol{d}(\boldsymbol{\xi})$, n is the sample size and $F(k, n-p\,; \alpha)$ is the upper $100(1-\alpha)$ percentile of the F-distribution with k and $n-p$ degrees of freedom. Del Castillo and Cahya (2001) provide a program, coded in the computer algebra package MAPLE, for the computation and display of the regions C_{BH}. Peterson, Cahya and del Castillo (2002) point out that the Box and Hunter procedure provides a confidence region for the *stationary* point, that is not necessarily a point of maximum. Therefore C_{BH} could consist in disconnected regions, because some points in C_{BH} could be associated with maximum points on the fitted response surface, while other points could be associated with saddle points. The same authors propose an alternative constrained approach, computing a confidence region for the optimum point of the response surface over an arbitrary experimental region and showing the connections with the Box-Hunter region.

This paper deals with the use of likelihood and Bayesian techniques to make inferences on the optimal factor combination. To make things easier, we introduce an alternative parametrization of M_P in which the true stationary point appears as parameter of interest. The structure of the paper is as follows. In Section 2 we describe the new proposed model. Section 3 derives the profile and integrated likelihoods for $\boldsymbol{\xi}$ and shows their behavior with a simulated example. In Section 4 we present the Bayesian analysis of the new model and an empirical method to obtain HPD regions for the true maximum point, which is obviously a special case of the stationary point. Three different examples of this method are given, based on experiments taken from the literature (Section 5, with $k = 3$ and Section 7, with $k = 5$) and on simulated data (Section 6, with $k = 2$). These examples show the HPD regions for the true maximum point that are obtained, through MCMC simulations, using non-informative and informative priors on $\boldsymbol{\xi}$. The C_{BH} confidence regions for the true stationary point are also computed to get a comparison. Finally, in Section 8 a simulation study is performed to compare the frequentist coverage probabilities of Bayesian and Box and Hunter regions and to examine the different widths of these regions.

2 The new proposed model

Since the interest is focused on the location of the true stationary point, we propose the use of an alternative model, equivalent to M_P and reparametrised so that $\boldsymbol{\xi}$ appears as parameter of interest.

The rationale for the polynomial approximation for the true response function is based on the Taylor series expansion around the center of the design. Assuming that a stationary point $\boldsymbol{\xi}$ for φ exists and it is unique, an alternative model can be obtained using $\boldsymbol{\xi}$ as origin for the Taylor expansion. We obtain this way the model

$$M_R: \quad \mathrm{y} = \alpha_0 + (\mathbf{x} - \boldsymbol{\xi})^T \mathbf{A} (\mathbf{x} - \boldsymbol{\xi}) + \varepsilon, \tag{3}$$

where $\mathbf{x}$ is a $k \times 1$ vector of factor levels, α_0 is the value of the true response function on $\boldsymbol{\xi}$ and $\mathbf{A}$ is the $k \times k$ symmetric matrix

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \frac{1}{2}\alpha_{12} & \cdots & \frac{1}{2}\alpha_{1k} \\ \frac{1}{2}\alpha_{12} & \alpha_{22} & \cdots & \frac{1}{2}\alpha_{2k} \\ \vdots & \vdots & & \vdots \\ \frac{1}{2}\alpha_{1k} & \frac{1}{2}\alpha_{2k} & \cdots & \alpha_{kk} \end{bmatrix}.$$

Model (3) is a one-to-one reparametrisation of the second order polynomial model: both models are characterized by $p + 1$ unknown parameters and the following relations between their parameters hold

$$\begin{cases} \beta_0 = \alpha_0 + \boldsymbol{\xi}^T \mathbf{A} \boldsymbol{\xi} \\ \boldsymbol{\beta} = -2\mathbf{A}\boldsymbol{\xi} \\ \mathbf{B} = \mathbf{A} \end{cases} \qquad\qquad \begin{cases} \alpha_0 = \beta_0 - \frac{1}{4}\boldsymbol{\beta}^T \mathbf{B}^{-1} \boldsymbol{\beta} \\ \boldsymbol{\xi} = -\frac{1}{2}\mathbf{B}^{-1}\boldsymbol{\beta} \\ \mathbf{A} = \mathbf{B} \end{cases}.$$

Therefore the MLEs of the parameters of M_R can be derived, through invariance, from the MLEs of the parameters of M_P, obtaining obviously that $\hat{\boldsymbol{\xi}}$ coincides with $\hat{\mathbf{x}}^S$. Such model is not useful in a frequentist framework because of the inconvenient nature of the sampling distribution of the estimated stationary point. On the contrary, it makes more direct and easier the likelihood techniques and the Bayesian approach in order to make inferences on $\boldsymbol{\xi}$, while the other parameters are treated as nuisance parameters.

3 Profile and integrated likelihoods

Let us suppose that n observations on the response variable are available. The model M_R can be written in the convenient matrix notation

$$M_R: \quad \mathbf{y} = \mathbf{X}_{\boldsymbol{\xi}} \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n), \tag{4}$$

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\boldsymbol{\alpha} = (\alpha_0, \alpha_{11}, ..., \alpha_{kk}, \alpha_{12}, ..., \alpha_{k-1,k})^T$ is the $p' \times 1$ vector of coefficients, with $p' = 1 + k + \frac{k(k-1)}{2}$, and $\boldsymbol{\varepsilon}$ is the vector of random errors. The $n \times p'$ matrix $\mathbf{X}_{\boldsymbol{\xi}}$ depends on the factor levels fixed by the experimenter and on the components of $\boldsymbol{\xi}$, the $k-$dimensional parameter of interest, in the following way

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

while $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \sigma^2)$ is the nuisance parameter of dimension $p' + 1$.

Denoting with $\mathbf{D}$ the matrix including the vector $\mathbf{y}$ and the fixed values of the k factors, the likelihood function for the complete vector of parameters is given by

$$\mathrm{L}(\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2 |\, \mathbf{D}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\boldsymbol{\alpha})^T(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\boldsymbol{\alpha})\Big\} \tag{5}$$

In order to eliminate the nuisance parameters, we can use standard likelihood techniques, such as profile and integrated likelihood. Replacing the nuisance parameters for each fixed $\boldsymbol{\xi}$ by their conditional MLEs, $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}} = (\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{X}_{\boldsymbol{\xi}})^{-1}\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{y}$ and $\hat{\sigma}^2_{\boldsymbol{\xi}} = \frac{1}{n}(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}})^T\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)$, we obtain the profile likelihood

$$\hat{\mathrm{L}}(\boldsymbol{\xi} |\, \mathbf{D}) = \mathrm{L}(\boldsymbol{\xi}, \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}, \hat{\sigma}^2_{\boldsymbol{\xi}} |\, \mathbf{D}) \propto \big[\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)^T\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)\big]^{-\frac{n}{2}}. \tag{6}$$

An alternative method is the elimination of nuisance parameters through integration, obtaining the so-called integrated likelihood

$$\tilde{\mathrm{L}}(\boldsymbol{\xi} |\, \mathbf{D}) \propto \int_{\Phi} \pi(\boldsymbol{\alpha}, \sigma^2 | \boldsymbol{\xi}) \mathrm{L}(\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{D}) \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\sigma^2,$$

where $\Phi = (0, \infty) \times (-\infty, \infty)^{p'}$ is the nuisance parameter space and $\pi(\boldsymbol{\alpha}, \sigma^2 | \boldsymbol{\xi})$ is a suitable weight function for $(\boldsymbol{\alpha}, \sigma^2)$ (see Berger, Liseo and Wolpert, 1999). In a Bayesian framework $\pi(\boldsymbol{\alpha}, \sigma^2 | \boldsymbol{\xi})$ represents the conditional prior distribution of $(\boldsymbol{\alpha}, \sigma^2)$ given the parameter of interest, $\boldsymbol{\xi}$. Let us remark that, given $\boldsymbol{\xi}$, model (4) has the structure of a normal linear model, with design matrix $\mathbf{X}_{\boldsymbol{\xi}}$. Therefore we can employ the usual non-informative (reference) prior distribution for the parameters of a normal linear model, that is $\pi^R(\boldsymbol{\alpha}, \sigma^2 | \boldsymbol{\xi}) \propto 1/\sigma^2$. Moreover the quadratic form in the exponent of (5) can be rewritten as

$$(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\boldsymbol{\alpha})^T(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\boldsymbol{\alpha}) = (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}})^T\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{X}_{\boldsymbol{\xi}}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}) + (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}})^T(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}})$$

so that

$$\begin{aligned} \tilde{\mathrm{L}}(\boldsymbol{\xi} |\, \mathbf{D}) &\propto \int_{\Phi} \frac{1}{\sigma^2} \mathrm{L}(\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{D}) \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\sigma^2 \\ &\propto \big|\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{X}_{\boldsymbol{\xi}}\big|^{-1/2} \int_0^{\infty} \frac{1}{(\sigma^2)^{\frac{n-p'+2}{2}}} \exp\Big\{ -\frac{(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}})^T(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}})}{2\sigma^2}\Big\} \mathrm{d}\sigma^2 \\ &\propto \big|\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{X}_{\boldsymbol{\xi}}\big|^{-1/2} \big[\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)^T\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)\big]^{-\frac{n-p'}{2}}, \end{aligned} \tag{7}$$

where $\big|\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{X}_{\boldsymbol{\xi}}\big|$ denotes the determinant of the matrix $\mathbf{X}_{\boldsymbol{\xi}}^T\mathbf{X}_{\boldsymbol{\xi}}$.

In this framework the parameter of interest $\boldsymbol{\xi}$ is *any possible type* of stationary point, not in particular the maximum. Given the data, we can explore not only the nature of the estimated stationary point, $\hat{\boldsymbol{\xi}}$, but also the nature of any possible value of $\boldsymbol{\xi}$. In fact, the estimated response surface, using model M_R, can be written for a given $\boldsymbol{\xi}$ as

$$\hat{\mathrm{y}}_{\boldsymbol{\xi}} = \hat{\alpha}_{0\boldsymbol{\xi}} + (\mathbf{x} - \boldsymbol{\xi})^T \hat{\mathbf{A}}_{\boldsymbol{\xi}} (\mathbf{x} - \boldsymbol{\xi}),$$

where

$$\hat{\mathbf{A}}_{\boldsymbol{\xi}} = \begin{bmatrix} \hat{\alpha}_{1\boldsymbol{\xi}} & \frac{1}{2}\hat{\alpha}_{12\boldsymbol{\xi}} & \cdots & \frac{1}{2}\hat{\alpha}_{1k\boldsymbol{\xi}} \\ \frac{1}{2}\hat{\alpha}_{12\boldsymbol{\xi}} & \hat{\alpha}_{2\boldsymbol{\xi}} & \cdots & \frac{1}{2}\hat{\alpha}_{2k\boldsymbol{\xi}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{2}\hat{\alpha}_{1k\boldsymbol{\xi}} & \frac{1}{2}\hat{\alpha}_{2k\boldsymbol{\xi}} & \cdots & \hat{\alpha}_{k\boldsymbol{\xi}} \end{bmatrix}. \tag{8}$$

Note that $\hat{\alpha}_{0\boldsymbol{\xi}}, \hat{\alpha}_{11\boldsymbol{\xi}}, ..., \hat{\alpha}_{kk\boldsymbol{\xi}}, \hat{\alpha}_{12\boldsymbol{\xi}}, ..., \hat{\alpha}_{k-1,k\boldsymbol{\xi}}$ are the elements of the conditional MLE $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}$. Now, the hessian matrix of $\hat{\mathrm{y}}_{\boldsymbol{\xi}}$, whose generic element is $h_{jh} = \left.\frac{\partial^2 \hat{\mathrm{y}}_{\boldsymbol{\xi}}}{\partial x_j x_h}\right|_{\mathbf{x}=\boldsymbol{\xi}}$, with $j, h = 1, ..., k$, is given by $2\hat{\mathbf{A}}_{\boldsymbol{\xi}}$. Therefore the experimental results partition the space Ξ of the $\boldsymbol{\xi}$ values into two sets, Ξ_{max} and $\Xi - \Xi_{max}$, where

$$\Xi_{max} = \{\boldsymbol{\xi} : \hat{\mathbf{A}}_{\boldsymbol{\xi}} \text{ is negative definite}\}$$

is the set of stationary points that, according to the data, turn out to be maximum points for $\hat{\mathrm{y}}_{\boldsymbol{\xi}}$. A careful exploration of the behavior of the profile and the integrated likelihoods requires taking into account this partition. The likelihoods (6) and (7) represent two different techniques for measuring the amount of support the data offer to the hypothesis that the point $\boldsymbol{\xi}$ under consideration is a stationary point. For every $\boldsymbol{\xi}$ it is possible to argue whether the support provided by the data is referred to a maximum point or not. Even if the estimated stationary point is a maximum, $\hat{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ and $\tilde{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ can assign a non-negligible support to the points $\boldsymbol{\xi} \in \Xi - \Xi_{Max}$, depending on the goodness of fit of the model and on how flat is the estimated surface around $\hat{\boldsymbol{\xi}}$. In general, if there is an adequate fit and the quadratic effects contribute significantly to the model, the points $\boldsymbol{\xi} \in \Xi - \Xi_{Max}$ should receive a negligible support. Next sub-section will illustrate the issue in more detail.

3.1 Two simulated examples with $k = 1$

The case $k = 1$ is not typical in a response surface study. However, this case is useful to show graphically the behavior of profile and integrated likelihoods.

Denoting with X the uncoded input variable, the estimated response surface for a given ξ is

$$\hat{\mathrm{y}}(\xi) = \hat{\alpha}_{0\xi} + \hat{\alpha}_{1\xi}(x - \xi)^2.$$

Given n experimental runs taken on the factor levels $x_1, ..., x_n$, the stationary points in Ξ_{Max} are all the values ξ such that

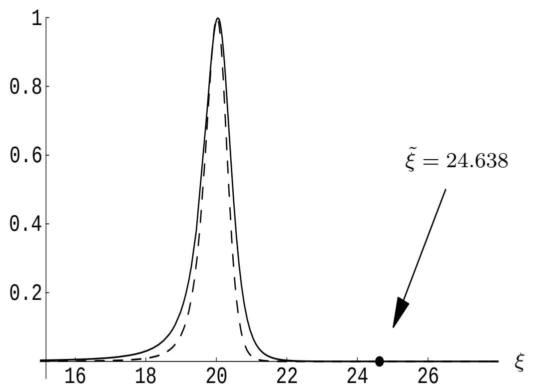
$$\hat{\alpha}_{1\xi} = \frac{\sum \mathrm{y}_i \sum (x_i - \xi)^2 - n \sum \mathrm{y}_i (x_i - \xi)^2}{\left[\sum (x_i - \xi)^2\right]^2 - n \sum (x_i - \xi)^4} < 0. \tag{9}$$

Therefore, in this simple situation, it is possible to obtain the threshold, say $\tilde{\xi}$, that divides Ξ into the two intervals Ξ_{max} and $\Xi - \Xi_{max}$. In fact, after some algebra, we obtain that the condition (9) is satisfied for all the values of ξ such that

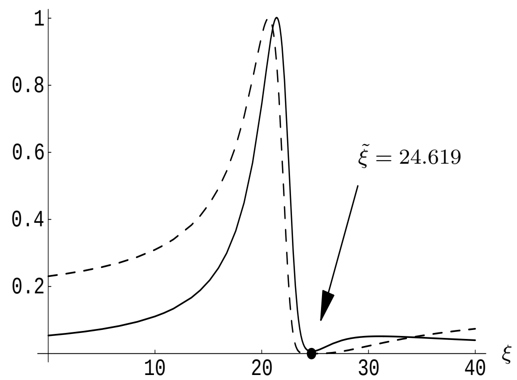
$$\begin{cases} \xi < \tilde{\xi} & \text{if } 2\big(\sum \mathrm{y}_i \sum x_i - n \sum \mathrm{y}_i x_i\big) > 0 \\ \xi > \tilde{\xi} & \text{if } 2\big(\sum \mathrm{y}_i \sum x_i - n \sum \mathrm{y}_i x_i\big) < 0 \end{cases}, \quad \text{where } \tilde{\xi} = \frac{\sum \mathrm{y}_i \sum x_i^2 - n \sum \mathrm{y}_i x_i^2}{2\big(\sum \mathrm{y}_i \sum x_i - n \sum \mathrm{y}_i x_i\big)}.$$

Two samples of size $n = 6$ were simulated from a true response function with maximum at $\xi = 20$, using the factor levels $\mathbf{x} = (22, 23, 24, 25, 26, 27)$ and standard errors simulated from a normal distribution of mean zero and variances equal to 1 and 2, respectively. The obtained data sets are $\mathbf{y_1} = (38.65, 35.35, 28.85, 22.20, 12.77, 2.74)$ and $\mathbf{y_2} = (38.16, 39.34, 31.14, 15.12, 11.93, -3.67)$. For data set $\mathbf{y_1}$, using the standard second order polynomial model, $\mathrm{R}^2 = 0.999$ indicates a very good fit of the quadratic polynomial model to the data and the quadratic term contributes significantly to the model (the $p-$value is 0.0016). For data set $\mathbf{y_2}$ we obtain $\mathrm{R}^2 = 0.962$ and the test for quadratic effect gives a $p-$value of 0.20. Therefore it is a typical situation in which further runs should be conducted. However, this example can be interesting to explore and to compare the behaviors of $\hat{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ and $\tilde{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ in such a situation.

Figure 1 shows the graphs of the profile and the integrated likelihood (normalized versions) for the two data sets. In both figures all the values of ξ such that $\xi < \tilde{\xi}$ represent the stationary points that are points of maximum for the fitted response. In particular, for the data set $\mathbf{y_2}$, being the quadratic effect nonsignificant, there is a non-negligible possibility that the true response is not a concave quadratic function. As a result, the profile and the integrated likelihood assign a positive support to the values of ξ which are not maximum. The different behavior of $\hat{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ and $\tilde{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ can be explained taking into account that the profile approach, replacing the nuisance parameters by their conditional MLEs, ignores their uncertainty. Integration methods, instead, "automatically incorporate nuisance parameter uncertainty, in the sense that an integrated likelihood is an average over all the possible conditional likelihoods given the nuisance parameter" (Berger, Liseo and Wolpert, 1999). On the contrary for data set $\mathbf{y_1}$ $\hat{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ and $\tilde{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D})$ almost coincide and the values of ξ which are not maxima, given the data, receive a zero support.



(a) Data set $\mathbf{y_1}$



(b) Data set $\mathbf{y_2}$

Figure 1: Profile (dashed lines) and integrated (full lines) likelihoods for ξ

4 Bayesian analysis of model M_R

Given a prior probability distribution for the parameters of M_R, $\pi(\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2) = \pi(\boldsymbol{\xi})\pi(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{\xi})$, the marginal posterior distribution of the parameter of interest is obtained by integrating out the nuisance parameters from the joint posterior, that is

$$\pi(\boldsymbol{\xi}|\mathbf{D}) \propto \pi(\boldsymbol{\xi}) \int_{\Phi} \pi(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{\xi}) \mathrm{L}(\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2|\mathbf{D}) \mathrm{d}\boldsymbol{\alpha} \mathrm{d}\sigma^2.$$

When clear prior information is lacking, the Bayesian method requires the use of non-informative prior distributions. Sambucini (2005), resorting to the well-known reference priors method (see, for instance, Bernardo, 1979; Berger and Bernardo, 1992a and 1992b) shows that the conditional reference prior of the nuisance parameters given the quantity of interest is $\pi^R(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{\xi}) \propto 1/\sigma^2$ and the marginal reference prior for the true stationary point is given by

$$\pi^R(\boldsymbol{\xi}) \propto \left|\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}}\right|^{-1/2}. \tag{10}$$

This is in general an improper distribution with its unique mode at the design center. Furthermore, $\pi^R(\boldsymbol{\xi})$ depends on the choice of the experimental design. When fitting a second order response surface, a standard property of the design is that of rotatability because the actual orientation of the system is generally unknown and rotatability assures that the sampling variance of the estimated response is constant for all the points at the same distance from the design center. Therefore, such points are treated as being "equally important". It is possible to show that, at least for standard cases, using a rotatable design the corresponding prior distribution (10) is constant for all the points at the same distance from the design center. These are then considered *a priori* as "equally probable" and a kind of rotational invariance is ensured from a Bayesian point of view as well.

If prior information about the location of the true stationary point is available, it is useful to introduce an informative marginal prior distribution for $\boldsymbol{\xi}$. Some possible choices are for instance the normal, the generalized Student-T and the uniform distribution on the experimental region, that is:

(i) $\pi^I(\boldsymbol{\xi}) = \mathrm{N}_k\big(\mu_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\xi}}\big), \ \forall \boldsymbol{\xi} \in \mathbb{R}^k;$

(ii) $\pi^I(\boldsymbol{\xi}) = \mathrm{T}_k\big(\mu_{\boldsymbol{\xi}}, \Sigma_{\boldsymbol{\xi}}; \nu\big), \ \forall \boldsymbol{\xi} \in \mathbb{R}^k;$

(iii) $\pi^I(\boldsymbol{\xi}) \propto c, \ \forall \boldsymbol{\xi} \in \mathcal{R}$ and c=constant.

The specification of the prior hyperperameters 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 $\mu_{\boldsymbol{\xi}} = \mathbf{0}$ and $\Sigma_{\boldsymbol{\xi}}$ diagonal with equal diagonal elements. In this way there is not a discrepancy between the knowledge the researcher expresses choosing the experimental design and that he wants to take into account introducing the informative prior distribution. The diagonal elements of $\Sigma_{\boldsymbol{\xi}}$ can be fixed with reference to the prior probability that seems reasonable to assign on $\mathcal{R}$.

Always assuming the conditional reference prior proportional to $1/\sigma^2$ for the nuisance parameters given $\boldsymbol{\xi}$, whether $\pi(\boldsymbol{\xi})$ is a non-informative prior or an informative one, the marginal posterior distribution of the true stationary point is

$$\begin{aligned} \pi(\boldsymbol{\xi}|\mathbf{D}) &\propto \pi(\boldsymbol{\xi})\, \tilde{\mathrm{L}}(\boldsymbol{\xi}|\,\mathbf{D}) \\ &= \pi(\boldsymbol{\xi}) \left|\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}}\right|^{-1/2} \left[\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)^T \big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)\right]^{-\frac{n-p'}{2}}. \end{aligned} \tag{11}$$

In particular, using the reference prior distributions, we obtain the reference marginal posterior distribution

$$\pi^R(\boldsymbol{\xi}|\mathbf{D}) \propto \left|\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}}\right|^{-1} \left[\big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)^T \big(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}} \hat{\boldsymbol{\alpha}}_{\boldsymbol{\xi}}\big)\right]^{-\frac{n-p'}{2}}. \tag{12}$$

If $k = 1$ cumbersome calculations show that (12) is a proper distribution. For a general k, heuristic arguments show that the posterior distribution, which can be written as a ratio of polynomials in $\boldsymbol{\xi}$, is again proper. For standard designs this can be easily proved at least for $k = 2, 3$.

4.1 An approximated method for posterior inferences on the true maximum point

It must be reminded that the parameter of interest is the stationary point for the true response function, not necessarily a maximum point. If we assume that a proper maximum point exists and we are interested in identifying it, we should impose *a priori* a constraint on the nuisance parameter space, assuming that matrix $\mathbf{A}$ in model M_R is negative definite. Then, denoting by Φ^* the subset of the nuisance parameter space Φ such that $\boldsymbol{\xi}$ is actually a maximum point, we should restrict our elaborations to the set Φ^*. Since this procedure is in general cumbersome, we propose a simpler approximated method.

The idea is to consider the marginal posterior distribution of the true stationary point and to identify the points $\boldsymbol{\xi} \in \Xi - \Xi_{max}$. Since, as previously remarked, these are all the stationary points that, given the data, are not of maximum response for the estimated surface, it is not reasonable to propose them as possible candidates for the maximum conditions. Therefore, to have a posterior density for the maximum point, we will restrict the posterior distribution of the stationary point to the set Ξ_{max}. The resulting density will be denoted with $\pi_{max}(\boldsymbol{\xi}|\mathbf{D})$. This procedure can be seen as an empirical technique, in which the parameter space Φ^* is actually estimated through the experimental results.

Now, our purpose is to make posterior inferences on the true maximum point, using its *approximated* marginal posterior distribution, $\pi_{max}(\boldsymbol{\xi}|\mathbf{D})$. In particular we are interested in computing highest posterior density (HPD) region. The $100(1-\alpha)\%$ HPD region for the true maximum point is given by

$$C_{HPD} = \left\{\boldsymbol{\xi} : \pi_{max}(\boldsymbol{\xi}|\mathbf{D}) \geq c_\alpha\right\}$$

where c_α is the largest number so that

$$1 - \alpha = \int_{\boldsymbol{\xi}:\, \pi_{max}(\boldsymbol{\xi}|\mathbf{D}) \geq c_\alpha} \pi_{max}(\boldsymbol{\xi}|\mathbf{D}) \mathrm{d}\boldsymbol{\xi}.$$

Note that $\pi_{max}(\boldsymbol{\xi}|\mathbf{D})$ can be computed up to a normalizing constant. Therefore we must employ Markov Chain Monte Carlo (MCMC) techniques. In particular in the following examples we shall use the Metropolis-Hastings algorithm, with a gaussian random walk to generate proposed values for the jump. Moreover, in order to be sure that we are drawing from $\pi^R_{max}(\boldsymbol{\xi}|\mathbf{D})$, in the algorithm we simply discard any point proposed for the jump that violate the constraint regarding matrix $\hat{\mathbf{A}}(\boldsymbol{\xi})$.

Assuming that an MCMC sample $\{\boldsymbol{\xi}_i,\, i = 1, ..., N\}$ is available from $\pi_{max}(\boldsymbol{\xi}|\mathbf{D})$, it is possible to estimate Bayesian HPD regions with the following steps (Chen, Shao and Ibrahim, 2000):

(i) compute $\zeta_i = \pi(\boldsymbol{\xi}_i|\mathbf{D})$, $\forall\, i = 1, ..., N$;

(ii) sort $\{\zeta_i,\, i = 1, ..., N\}$ to obtain the ordered values $\zeta_{(1)} \leq \zeta_{(2)} \leq ... \leq \zeta_{(N)}$;

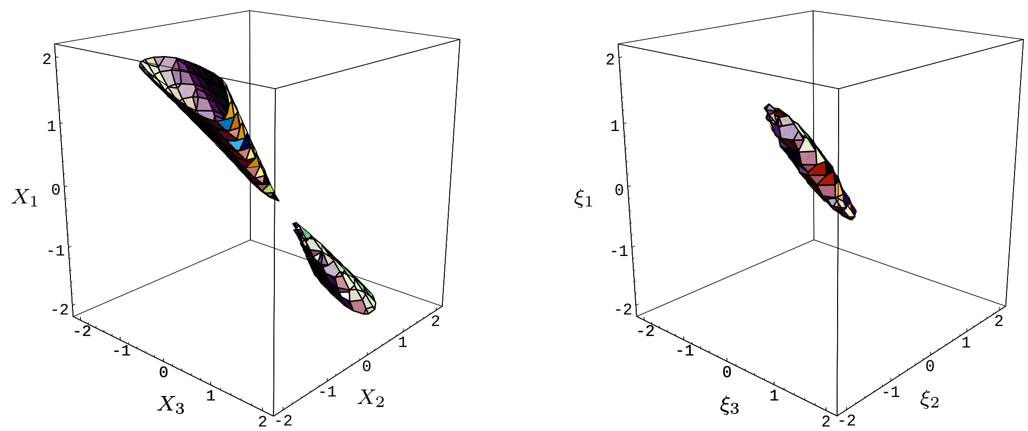
(iii) compute the $100(1-\alpha)\%$ HPD region for $\boldsymbol{\xi}$, using $c_\alpha = \zeta_{(\alpha N)}$.

When $k = 2, 3$ it is possible to plot the C_{HPD} regions displaying in a 2−dimensional or 3−dimensional space, respectively, the set of points $\boldsymbol{\xi}$ such that $\pi_{max}(\boldsymbol{\xi}|\mathbf{D}) = c_\alpha$, using the value of c_α estimated through MCMC simulations. For $k > 3$ we can show the point-by-point projections of the k−dimensional HPD region into 2−dimensional planes. Given an MCMC sample $\{\boldsymbol{\xi}_i,\ i = 1, ..., N\}$ from $\pi_{max}(\boldsymbol{\xi}|\mathbf{D})$, for each point $\boldsymbol{\xi}^* = (\xi_1^*, ..., \xi_i^*, ..., \xi_j^*, ..., \xi_k^*)$ in the sample that lies inside the C_{HPD} region the idea is to plot the points (ξ_i^*, ξ_j^*) on the $i - j$ planes, $\forall\, i, j = 1, ..., k,\ i < j$. Such point-by-point projections represent the "shades" of the k−dimensional HPD regions on the $\frac{k(k-1)}{2}$ possible 2−dimensional spaces. Of course the points inside the projections actually lie inside the k−dimensional C_{HPD} only for some values of the other $k - 2$ factors.

In the examples which follow the HPD regions will be compared with Box and Hunter confidence regions referring to the same levels of probability and confidence (usually 90% and 95%) respectively.

5 A three factor experiment

Let us consider the experiment described in Section 9.2 of Box and Draper (1987) and also studied by del Castillo and Cahya (2001) and Peterson, Cahya and del Castillo (2002). In particular, these latter authors used this example to show the possibility of getting disconnected Box and Hunter confidence regions for the stationary point. In the experiment the purpose is to find the combination of the levels of three reaction conditions that maximize the elasticity of a certain polymer. The factors of interest, denoted as coded variables with X_1, X_2 and X_3, are the percentage concentrations of two constituents and the reaction temperature. A central composite design is employed, with two center runs and six axial points $(\pm 2, 0, 0)$, $(0, \pm 2, 0)$ and $(0, 0, \pm 2)$. Fitting the second order polynomial model, all the terms in the model are statistically significant and the R^2 value is 0.972 ($R^2_{adj} = 0.930$), showing an adequate fit. The estimated stationary point, $\hat{\mathbf{x}}^S = (0.460, -0.464, 0.151)$, located inside the experimental region, turns out to be a maximum point being all the eigenvalues of $\hat{\mathbf{B}}$ negative.



(a) 95% C_{BH} (b) 95% C_{HPD}, using $\pi^R(\boldsymbol{\xi})$

Figure 2: Regions C_{BH} for the true stationary point and C_{HPD} for the true maximum point.

To perform a Bayesian inference, we employ the Random Walk Metropolis-Hastings algorithm to generate a sample of simulations from the approximated marginal posterior distribution of the maximum point, obtained through the reference prior (10). In particular we draw a sample of size 70000, discarding the initial 20000 iterations and taking every 10th simulation draw. The result is a simulated chain of size 5000.

Figure 2 (a) and (b) shows the 95% Box and Hunter confidence region for the true stationary point and the 95% HPD approximated region for the true maximum point. The C_{BH} region is composed of two disjoint regions, one associated with maximum points and the other associated with saddle points (as already observed by Peterson, Cahya and del Castillo, 2002). Furthermore, the region is unbounded and therefore it is not contained into the experimental region. On the contrary, C_{HPD} is properly contained in the experimental region.

6 Simulated examples with $k = 2$

Simulated examples, differently from real data examples, allow to show whether the calculated regions contain or not the true maximum point. Different patterns are exhibited in the examples which follow.

Three samples of size $n = 12$ are simulated from the coded true response function

$$\varphi(x_1, x_2) \quad = \quad 100 - 2.5(x_1 - 0.6)^2 - 5(x_2 - 0.4)^2 + 3.75(x_1 - 0.6)(x_2 - 0.4), \tag{13}$$

with a maximum at $\boldsymbol{\xi} = (0.6, 0.4)$, using simulated errors from a normal distribution of mean zero and σ^2 equal to 1, 2 and 3, respectively. The combinations of factor levels considered are those of a rotatable central composite design with axial points at distance $\sqrt{2}$ from the design center and 4 center points. The obtained data sets are in Table 1. For the three data sets, fitting the second order polynomial model, the quadratic effects are statistically significant and the R^2 values are, respectively, 0.982, 0.925 and 0.921. The estimated stationary points are

$$\hat{\boldsymbol{\xi}}_{\mathbf{y_1}} = (0.563, 0.371), \qquad \hat{\boldsymbol{\xi}}_{\mathbf{y_2}} = (0.315, 0.235) \qquad \text{and} \qquad \hat{\boldsymbol{\xi}}_{\mathbf{y_3}} = (0.465, 0.333)$$

and they correspond to maximum points that lie inside the experimental region.

Table 1: Simulated runs from (13), using $\sigma^2 = 1$, $\sigma^2 = 2$ and $\sigma^2 = 3$.

Runs	X_1	X_2	$\mathbf{y_1}$	$\mathbf{y_2}$	$\mathbf{y_3}$
1	-1	-1	90.11	94.64	93.53
2	-1	1	88.01	88.26	87.07
3	1	-1	88.67	89.64	85.85
4	1	1	98.58	95.65	97.21
5	0	0	99.10	99.93	96.05
6	0	0	99.48	99.01	102.32
7	0	0	99.01	99.90	97.74
8	0	0	100.24	99.39	98.91
9	$-\sqrt{2}$	0	92.13	92.09	90.27
10	$\sqrt{2}$	0	99.98	95.66	94.68
11	0	$-\sqrt{2}$	88.25	89.16	82.99
12	0	$\sqrt{2}$	93.00	93.99	89.73

Figure 3 shows the C_{BH} regions for the true stationary point and the approximated HPD regions for the true maximum point, obtained through MCMC simulations introducing the following marginal prior distributions for $\boldsymbol{\xi}$:

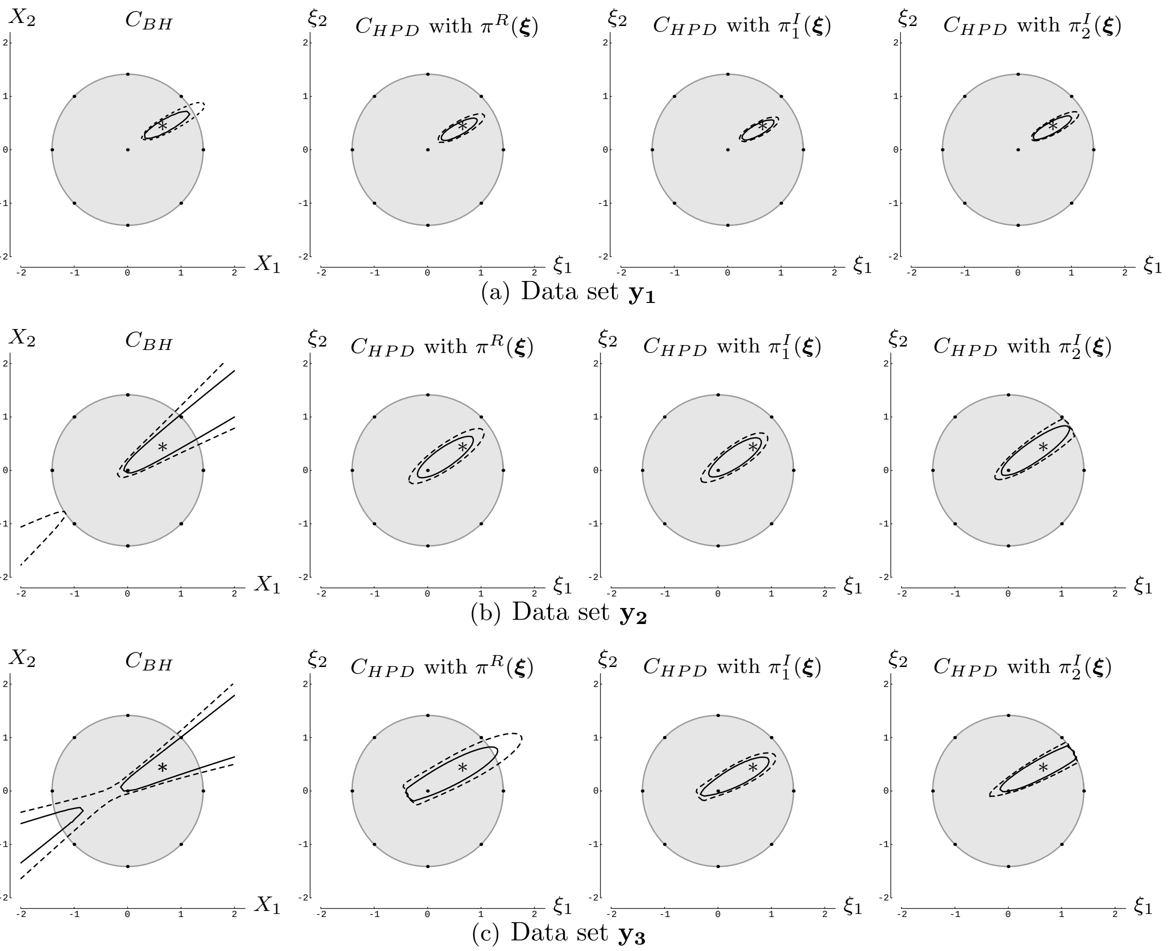
(i) the non-informative reference prior, $\pi^R(\boldsymbol{\xi})$;

(ii) the informative normal distribution $\pi_1^I(\boldsymbol{\xi}) = N_2(\mathbf{0}, \Sigma_{\boldsymbol{\xi}})$, with $\Sigma_{\boldsymbol{\xi}} = \text{diag}(0.33, 0.33)$;

(iii) the informative uniform distribution in the experimental region, $\pi_2^I(\boldsymbol{\xi})$.

Note that the informative priors do not express information about the true orientation of the system and therefore they are coherent with the design property of rotatability. It is sensible that in general the informative priors assign an high probability to the experimental region $\mathcal{R}$. In particular, under $\pi_1^I(\boldsymbol{\xi})$, the region $\mathcal{R}$ has a probability approximately equal to 0.95.

All the regions contain the true maximum point, but there is an apparent difference between the sizes of the regions. The already encountered situations of disconnectedness of C_{BH} regions again occur in some cases.



(a) Data set $\mathbf{y_1}$

(b) Data set $\mathbf{y_2}$

(c) Data set $\mathbf{y_3}$

Figure 3: 90% (full lines) and 95% (dashed lines) C_{BH} for the true stationary point and C_{HPD} for the true maximum point. The gray areas represent the experimental region. The experimental runs are denoted by • and the true maximum point by $*$.

7 A five factor experiment

The following example is considered several times in the literature (Box, 1954; Box and Draper (1987); Bisgaard and Ankenman, 1996; del Castillo and Cahya, 2001), which should be referred for more details. It involves the maximization of the yield of a chemical process with two stages and includes five factors: the temperatures (X_1 and X_4) and the reaction times (X_2 and X_5) at the two stages and the concentration of one of the reactants at the first stage. Fitting the second order polynomial model on the basis of 32 runs the estimated stationary point is $\hat{\mathbf{x}}^S = (2.50, -1.09, 1.24, -0.30, 0.54)$ and represents a point of maximum.

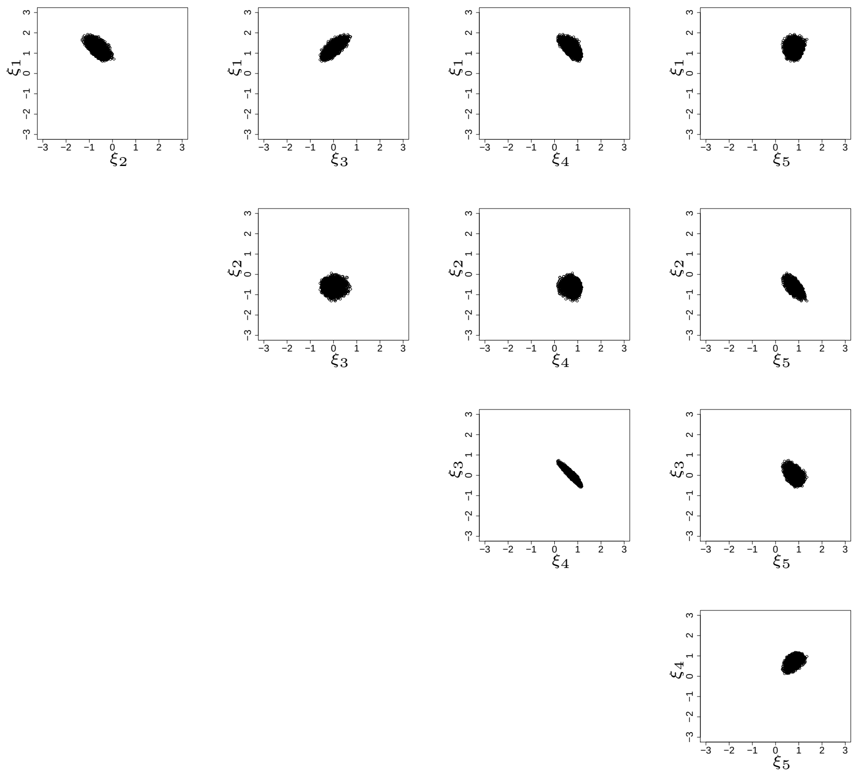


Figure 4: Point-by-point projections of the 90% C_{HPD} region for the true maximum point, using the non-informative reference prior for $\boldsymbol{\xi}$.

Figure 4 shows the point-by-point projections of the 5-dimensional 90% HPD region for the true maximum point obtained using the reference prior (10) for $\boldsymbol{\xi}$. The figure is obtained using MCMC simulations as described in Section 4.1. In order to make a comparison, we also report Figure 5, taken from del Castillo and Cahya (2001), in which the point-by-point projections of the 90% Box and Hunter confidence region for the true stationary point are displayed. These authors obtained the figure computing a 5-dimensional grid of points and plotting on the $i - j$ plane every sub-vector (x_i, x_j) which is a component of a 5-dimensional vector in the grid such that it lies in the C_{BH} region for at least one choice of the other three factors levels. The projections of the 90% C_{BH} region are clearly larger than the projections of the 90% HPD region.

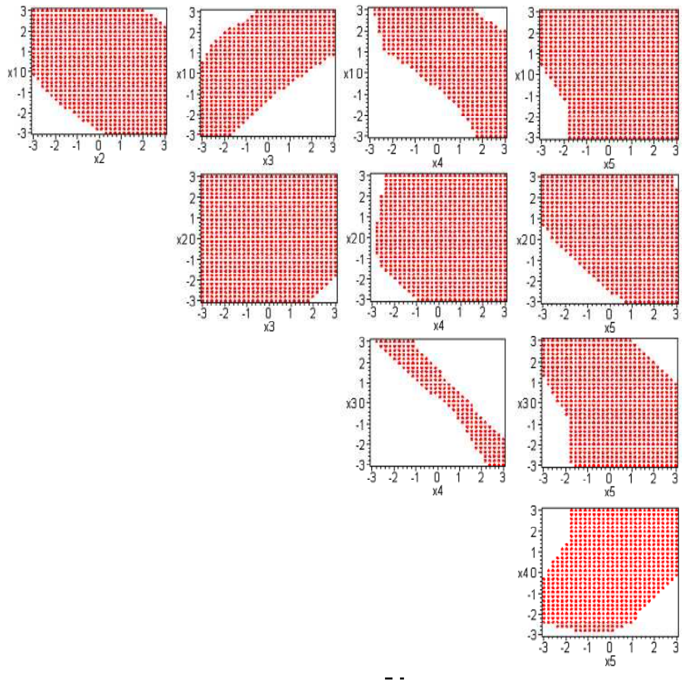


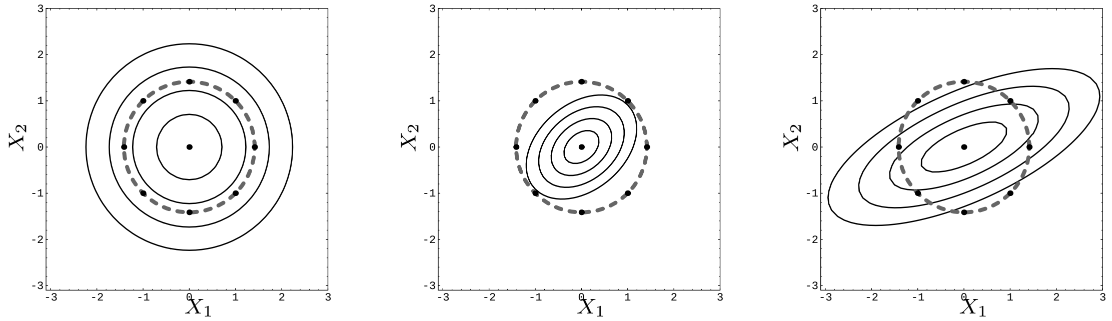
Figure 5: Point-by-point projections of the 90% C_{BH} region for the true stationary point (del Castillo and Cahya, 2001).

8 Coverage rates

In order to assess if the proposed Bayesian procedures have a good behavior from a frequentist perspective, we provide a simulation study to obtain an empirical check on the coverage probability of the Bayesian posterior regions. The coverage probability of the Box and Hunter confidence regions is also evaluated to have a comparison. In addition to the coverage rates it is interesting to examine the width of the regions to assess if good coverage rates are associated with big sizes, hence with poorly informative conclusions.

Given two coded factors, we consider different true response functions (modeled as in (3)) with a unique maximum point. From each of them, 1000 samples of $n = 12$ response values are simulated corresponding to the experimental points of a rotatable central composite design with four center points. The experimental region, $\mathcal{R}$, is a circle of radius $\sqrt{2}$. The simulated errors are from a normal distribution with mean zero and variance 1. The true maximum point has been located on the design center ($\boldsymbol{\xi} = (0,0)$), inside $\mathcal{R}$ ($\boldsymbol{\xi} = (0.5, 0.5)$), on the boundary of $\mathcal{R}$ ($\boldsymbol{\xi} = (1,1)$) and then outside of the experimental region ($\boldsymbol{\xi} = (1.5, 1.5)$). The value of the true nuisance parameter $\boldsymbol{\alpha}$ is also varied to consider different shapes and curvatures. Choices of $\boldsymbol{\alpha}$ which lead to high eigenvalues of matrix $\mathbf{A}$ characterize a more peaked true response function, while a considerable difference in the magnitude of this eigenvalues indicates an elongation in the surface. Equal eigenvalues results in a rotationally symmetric true response function. Figure 6 shows the three true surfaces considered in the simulation study for the three chosen values of $\boldsymbol{\alpha}$ when $\boldsymbol{\xi} = (0,0)$ (for the other values of $\boldsymbol{\xi}$ the surfaces are rigidly translated).

The simulated data are used to compute the coverage rates of the 95% Box and Hunter confidence regions for the true stationary point and of the 95% Bayesian HPD regions for the true maximum point. For the Bayesian regions we consider three different marginal prior distribution for $\boldsymbol{\xi}$: the non-informative reference prior (10), an informative normal distribution with mean $\mathbf{0}$



(a) $\boldsymbol{\alpha} = (100, -2, -2, 0)$ (b) $\boldsymbol{\alpha} = (100, -8, -9, 6)$ (c) $\boldsymbol{\alpha} = (100, -2, -6, 4.5)$

Figure 6: Contours of levels 99, 97, 94 and 90 of the true response functions considered in the simulation study for the three chosen values of $\boldsymbol{\alpha}$ when $\boldsymbol{\xi} = (0, 0)$. The true maximum response is 100, being $\alpha_0 = 100$. The experimental runs are denoted by • and the dashed circles represent the experimental region.

and variance-covariance matrix $\Sigma_{\boldsymbol{\xi}} = \text{diag}(0.621, 0.621)$ and an informative uniform distribution on $\mathcal{R}$. They are respectively denoted with $\pi^R(\boldsymbol{\xi})$, $\pi^N(\boldsymbol{\xi})$ and $\pi^U(\boldsymbol{\xi})$. Note that the informative priors are coherent with the choice of a rotatable design. Moreover, the informative distribution $\pi^N(\boldsymbol{\xi})$ assigns a prior probability to the experimental region approximately equal to 0.8. Let us remark that we can evaluate the coverage rates of the HPD regions only through MCMC methods: thus this check requires long computational times. Besides the coverage probability, in order to take into account the width of the C_{BH} and C_{HPD} regions, we compute the proportion of regions that cover the true maximum and are completely enclosed in $\mathcal{R}$ or in the regions $\mathcal{R}_2$ and $\mathcal{R}_3$, which are circles with double and triple radius with respect to the experimental region.

The simulation results are given in Table 2. In most cases the observed coverage probabilities of the HPD regions are reasonably close to the nominal value, showing a good frequentist behavior of the Bayesian procedure in particular when the non-informative reference prior for $\boldsymbol{\xi}$ is assumed. When we use the informative distribution $\pi^N(\boldsymbol{\xi})$, the coverage rates tend to be lower than the nominal level as the true maximum point get far the design center mainly in the case $\boldsymbol{\alpha} = (100, -2, -6, 4.5)$, when the true surface is more flat and more elongated. This fact is not surprising. The informative distribution $\pi^N(\boldsymbol{\xi})$, being a normal with zero mean, indicates *a priori* the design center as the best candidate for the maximum conditions and assigns a prior probability to the experimental region approximately equal to 0.8. On the other hand, when the true maximum point is located on the boundary or outside of the experimental region, there is a big percentage of simulated data sets with an estimated stationary point that lies outside of $\mathcal{R}$. In particular, for the three values of $\boldsymbol{\alpha}$ considered, these percentages are respectively 53%, 50.1% and 53.5% when $\boldsymbol{\xi} = (1, 1)$ and 99.4%, 100% and 92.8% when $\boldsymbol{\xi} = (1.5, 1.5)$. In such situations, according to the typical RSM procedures (see, for instance, Cornell, 1990), it is reasonable to conduct a further experiment rather than to construct confidence regions for $\boldsymbol{\xi}$. This is the reason for which in this simulation study we do not consider true response functions with maximum point extremely far from $\mathcal{R}$.

As regards the widths of the C_{BH} and C_{HPD} regions, let us remark that in all cases the proportion of HPD regions that cover the true maximum and are completely enclosed in $\mathcal{R}$, $\mathcal{R}_2$ or $\mathcal{R}_3$ is always larger than the corresponding proportion of the Box and Hunter confidence regions.

Table 2: Coverage rates for 95% C_{BH} and C_{HPD} regions and proportions of regions that cover the true $\boldsymbol{\xi}$ and are inside $\mathcal{R}$, $\mathcal{R}_2$ or $\mathcal{R}_3$, for different values of $\boldsymbol{\xi}$ and $\boldsymbol{\alpha} = (\alpha_0, \alpha_{11}, \alpha_{22}, \alpha_{12})$.

		$\boldsymbol{\alpha} = (100, -2, -2, 0)$ Eigenvalues $\mathbf{A} = (-2,\ -2)$				$\boldsymbol{\alpha} = (100, -8, -9, 6)$ Eigenvalues $\mathbf{A} = (-11.54,\ -5.46)$				$\boldsymbol{\alpha} = (100, -2, -6, 4.5)$ Eigenvalues $\mathbf{A} = (-7.01,\ -0.99)$			
		C_{BH}	C_{HPD}			C_{BH}	C_{HPD}			C_{BH}	C_{HPD}		
$\boldsymbol{\xi}$			$\pi^R(\boldsymbol{\xi})$	$\pi^N(\boldsymbol{\xi})$	$\pi^U(\boldsymbol{\xi})$		$\pi^R(\boldsymbol{\xi})$	$\pi^N(\boldsymbol{\xi})$	$\pi^U(\boldsymbol{\xi})$		$\pi^R(\boldsymbol{\xi})$	$\pi^N(\boldsymbol{\xi})$	$\pi^U(\boldsymbol{\xi})$
$(0, 0)$	Coverage	0.957	0.978	0.975	0.978	0.945	0.945	0.952	0.947	0.945	0.968	0.965	0.959
	$\in \mathcal{R}$	0.734	0.965	0.987	–	1	1	1	–	0.251	0.777	0.841	–
$(0.5, 0.5)$	Coverage	0.961	0.974	0.975	0.964	0.939	0.939	0.943	0.941	0.951	0.960	0.959	0.958
	$\in \mathcal{R}$	0.245	0.451	0.560	–	0.995	1	1	–	0.056	0.245	0.308	–
	$\in \mathcal{R}_2$	0.586	0.946	1	–	0.995	1	1	–	0.160	0.791	0.997	–
$(1, 1)$	Coverage	0.961	0.965	0.952	–	0.942	0.939	0.936	–	0.945	0.954	0.930	–
	$\in \mathcal{R}_2$	0.271	0.564	0.991	–	0.997	1	1	–	0.055	0.356	0.973	–
	$\in \mathcal{R}_3$	0.475	0.867	1	–	1	1	1	–	0.112	0.657	1	–
$(1.5, 1.5)$	Coverage	0.957	0.960	0.893	–	0.940	0.937	0.926	–	0.950	0.950	0.786	–
	$\in \mathcal{R}_2$	0.053	0.124	0.711	–	0.607	0.750	0.962	–	0.017	0.096	0.775	–
	$\in \mathcal{R}_3$	0.274	0.569	1	–	0.998	1	1	–	0.072	0.398	1	–

Acknowledgements

The authors wish to thank Marco Perone Pacifico for helpful suggestions.

References

Berger, J.O. and Bernardo, J.M. (1992a). Ordered group reference priors with application to a multinomial problem. *Biometrika*, **79**, 25-37.

Berger, J.O. and Bernardo, J.M. (1992b). On the development of reference priors (with discussion). In *Bayesian Statistics 4*, Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, 35-60. Oxford University Press, Oxford.

Berger, J.O., Liseo, B. and Wolpert, R.L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, Vol. **14**, No. 1, 1-28.

Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society*, **41**, 113-147 (with discussion).

Bisgaard, S. and Ankenman, B. (1996): Standard errors for the eigenvalues in second-order response surface models, *Technometrics*, **38**, No 3, 238-246.

Box, G.E.P. (1954). The exploration and exploitation of response surfaces: some general considerations and examples. *Biometrics*, **10**, 16-60.

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., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for experiments*, Wiley, New York.

Box, G.E.P. and Wilson, K.B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society*, **B**, **13**, 1-45.

Carter, W.H., Chinchilli, V.M., Myers, R.H. and Campbell, E.D. (1986): Confidence intervals and an improved ridge analysis of response surfaces, *Technometrics*, **28**, 339-346.

Carter, W.H., Chinchilli, V.M. and Campbell, E.D. (1990): A large-sample confidence region useful in characterizing the stationary point of a quadratic response surface, *Technometrics*, **32**, 425-435.

Chen, M.-H, Shao, Q.-M. and Ibrahim, J.G (2000). *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York.

Cornell, J.A. (1990). *How to apply Response surface methodology*, American Society for Quality Control - Statistics Division, ASQC, Milwaukee.

Davies, O.L. (1960). *The Design and Analysis of Industrial Experiments* Hafner Publishing Company, New York.

Del Castillo, E. and Cahya, S. (2001): A tool for computing confidence regions on the stationary point of a response surface, *The American Statistician*, **55** , 4, 358-365.

Draper, N.R. (1963): Ridge analysis of response surfaces. *Technometrics*, **5**, 469-479.

Graybill, F.A. (1969): *Introduction to Matrices with Applications in Statistics*, Wadsworth Publishing Company, Belmont.

Khuri, A.I. and Cornell, J.A. (1987). *Response Surfaces: Design and Analyses*, Marcel Dekker, New York.

Myers, R.H. and Montgomery, D.C. (1995). *Response Surface Methodology: Process and Produce Optimization Using Designed Experiments*, Wiley, New York.

Peterson, J.J. (1993). A general approach to ridge analysis with confidence intervals. *Technometrics*, **35**, 204-214.

Peterson, J.J., Cahya, S. and del Castillo, E. (2002). A general approach to confidence regions for optimal factor levels of response surfaces. *Biometrics*, **58**, 422-431.

Sambucini, V. (2005). A reference prior for the analysis of a response surface. Technical Report n.15, Department of Statistics, Probability and Applied Statistics, University of Rome "La Sapienza", Italy.