A reference prior for the analysis of a response surface

Valeria Sambucini
Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università "La Sapienza", Piazzale Aldo Moro 5, Roma 00185, Italia
valeria.sambucini@uniroma1.it

Abstract

Standard response surface methodology employs a second order polynomial model to locate the stationary point $\boldsymbol{\xi}$ of the true response function. To make Bayesian analysis more direct and simpler, we refer to an alternative and equivalent parametrization, which contains $\boldsymbol{\xi}$ as parameter of interest. The marginal reference prior of $\boldsymbol{\xi}$ is derived in its general form and particular cases are also given in detail, showing the Bayesian role of rotatability.

Keywords: Bayesian analysis, central composite designs, factorial designs, reference priors, response surface methodology, rotatability.

1 Introduction

Given a response variable Y and k continuous factors, $U_1, ..., U_k$, the main purpose of Response Surface Methodology (RSM), introduced by Box and Wilson (1951), is to find the combination of factor levels to achieve the optimal response. The typical RSM procedures are described in detail in many textbooks, including those by Davies (1960), Box and Draper (1987), Khuri and Cornell (1987) and Myers and Montgomery (1995).

For computational convenience, the natural variables are usually converted to coded or design variables, $X_1, ..., X_k$, standardized so that the design center is at the point $(x_1, ..., x_k) = \mathbf{0}$. Moreover it is assumed that the true response is a function of the levels of the k design variables, $\varphi(x_1, x_2, ..., x_k)$, called the *true response function*. The goal of the analysis is to investigate the behavior of the unknown function φ over the *operability region*, $\mathcal{R}_O$, that is the whole subregion of $\mathbb{R}^k$ where it is theoretically possible to do the experiment and observe response values. To make it possible, the researcher usually defines a smaller subregion $\mathcal{R} \subset \mathcal{R}_O$, called *experimental region* or *region of interest*, that is the region over which the experiment will be performed.

Since the actual form of φ is generally unknown, it is usually approximated by a polynomial of first or second degree in the confined region $\mathcal{R}$. A second order polynomial is used in particular when the interest is focused on the location of the stationary point of the true response function, while for the sequential exploration of the surface a first degree polynomial can be sufficient.

The second order polynomial model can be compactly written as

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

where ε is the random error which is assumed to be distributed as a normal distribution 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, $\boldsymbol{b}$ 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 $(1/2)\beta_{ij}$. From the fitted second order surface model, the estimated stationary point, $\hat{\mathbf{x}}^S = -\frac{1}{2}\hat{\mathbf{B}}^{-1}\hat{\boldsymbol{b}}$, is computed, where $\hat{\boldsymbol{b}}$ and $\hat{\mathbf{B}}$ are the maximum likelihood estimates (MLEs) of $\boldsymbol{b}$ and $\mathbf{B}$. Depending on the eigenvalues of $\hat{\mathbf{B}}$, $\hat{\mathbf{x}}^S$ could be a minimum, a maximum or a saddle point. Therefore a "canonical analysis" is 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, confidence regions on the location of the true stationary point are derived (Box and Hunter, 1954) to get a more realistic assessment of the quality of the point estimate.

Formula (1) can be justified by the use of the Taylor formula with the origin at the center of the design. Assuming that a stationary point $\boldsymbol{\xi}$ for φ exists and it is unique, an alternative (and equivalent) parametrization can be obtained using $\boldsymbol{\xi}$ as origin for the Taylor expansion. We obtain this way the model

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

where 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 $(1/2)\alpha_{ij}$. Both models, (1) and (2), are characterized by $p+1$ unknown parameters, where $p = 1 + 2k + \frac{k(k-1)}{2}$.

The relations between the parameters of the two models are

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

and the MLEs of the parameters of M_R can be derived, through invariance, from the MLEs of the parameters of M_P. Obviously $\hat{\boldsymbol{\xi}} = -\frac{1}{2}\hat{\mathbf{B}}^{-1}\hat{\boldsymbol{b}}$ coincides with $\hat{\mathbf{x}}^S$. Introducing $\boldsymbol{\xi}$ as a parameter would make the standard frequentist elaborations quite difficult, due to the nature of the sampling distribution of its MLE, $\hat{\boldsymbol{\xi}}$. In fact this suggestion does not appear in literature, but a Bayesian analysis becomes more direct and even simpler since the inferential interest is only centered on $\boldsymbol{\xi}$.

In order to perform a Bayesian analysis a prior probability distribution for the parameters must be introduced. This can represent actual pre-experimental information about the response surface or we can resort to some kind of non-informative prior. The most used and general technique is the so-called reference prior method. It was proposed by Bernardo (1979), whose original idea was to find the prior distribution that maximizes the missing information about the quantity of interest. Since Bernardo's original paper, the reference prior method has been developed and refined (see, for instance, Berger and Bernardo, 1989, 1992a, 1992b, and 1992c) and also applied to various problems (Bayarri, 1981; Ye and Berger, 1991; Liseo, 1993; Sun and Ye, 1995; Bernardo and Ramón, 1998; to name just a few). Since most of these papers are due to Berger and Bernardo, the statistical literature often refers to the reference prior method as the Berger-Bernardo method.

An important innovation in constructing non-informative priors which characterizes the Berger-Bernardo method is the different treatment of interest and nuisance parameters. In multiparameter situations the general $m-$group reference prior algorithm requires dividing the whole vector of parameters, $\boldsymbol{\theta}$, into m ordered groups $\boldsymbol{\theta}_{(1)}, ..., \boldsymbol{\theta}_{(m)}$, where each group $\boldsymbol{\theta}_{(j)}$ contains one or more of the scalar parameters in $\boldsymbol{\theta}$. It is suggested the ordering be in terms of inferential importance of the groups: the parameter of interest should be the first. Ignoring some technical problems, the iterative procedure is as follows: first find the conditional reference prior for $\boldsymbol{\theta}_{(m)}$ given $(\boldsymbol{\theta}_{(1)}, ..., \boldsymbol{\theta}_{(m-1)})$, then find the conditional reference prior for $\boldsymbol{\theta}_{(m-1)}$ given $(\boldsymbol{\theta}_{(1)}, ..., \boldsymbol{\theta}_{(m-2)})$ and so on, until computing the marginal reference prior for the parameters of interest, $\boldsymbol{\theta}_{(1)}$. The conditional reference prior $\pi(\boldsymbol{\theta}_{(j)}|\boldsymbol{\theta}_{(1)}, ..., \boldsymbol{\theta}_{(j-1)})$, $\forall j = 1, ..., m-1$, is based on the marginal model obtained by integrating

out the parameters $\boldsymbol{\theta}_{(j+1)}, ..., \boldsymbol{\theta}_{(m)}$. The reference priors obtained can depend on the grouping. Berger and Bernardo (1992c) recommend using the one-at-a-time reference prior, that is to have one parameter per group, "unless there is a specific reason for using a certain grouping". The algorithm is typically very hard to implement, but in the regular case, where joint posterior asymptotic normality may be established, calculations become considerably simpler.

Section 2 shows a reference analysis of model M_R. In this application it is sensible to refer to the parameter of interest as the whole vector of the coordinates of the true stationary point. In Section 3, when $k = 2$, some features of the marginal reference prior of $\boldsymbol{\xi}$ are described, depending on the properties of the experimental design.

2 A reference analysis of model M_R

2.1 Reference prior distribution of $\boldsymbol{\xi}$

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

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

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_k, \alpha_{12}, ..., \alpha_{k-1,k})^T$ is the $p' \times 1$ vector of coefficients, $\boldsymbol{\varepsilon}$ is the vector of random errors and

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

Here $\boldsymbol{\xi}$ is the $k-$dimensional parameter of interest and $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \sigma^2)$ is the nuisance parameter of dimension $p' + 1$, where $p' = 1 + k + \frac{k(k-1)}{2}$. Note that, given $\boldsymbol{\xi}$, model (3) has the structure of a normal linear model, with design matrix $\mathbf{X}_{\boldsymbol{\xi}}$. Therefore the conditional reference prior of the nuisance parameter given the quantity of interest is the standard one

$$\pi^R(\boldsymbol{\alpha}, \sigma^2 | \boldsymbol{\xi}) \propto \frac{1}{\sigma^2}. \tag{4}$$

Since M_R is a one-to-one reparametrisation of the second order polynomial model, we are in the regular case in which asymptotic normality of the posterior holds (see Bernardo and Smith, 1994, Section 5.3.3). Let us denote with $\mathbf{S}_{\xi\xi}(\boldsymbol{\xi}, \boldsymbol{\phi})$ the $(k \times k)$ upper matrix of $\mathbf{S}_{\boldsymbol{\psi}}(\boldsymbol{\xi}, \boldsymbol{\phi}) = \mathbf{I}^{-1}_{\boldsymbol{\psi}}(\boldsymbol{\xi}, \boldsymbol{\phi})$, where $\mathbf{I}_{\boldsymbol{\psi}}(\boldsymbol{\xi}, \boldsymbol{\phi})$ is the Fisher Information matrix. Under posterior asymptotic normality, we can generalize Proposition 3 and its Corollary of Bernardo and Ramón (1998) to the case of a parameter of interest which is itself a vector. Then if the nuisance parameter space is independent of the parameter of interest and the following factorization holds

$$\left|\mathbf{S}^{-1}_{\xi\xi}(\boldsymbol{\xi}, \boldsymbol{\phi})\right|^{1/2} = f_0(\boldsymbol{\xi}) g_0(\boldsymbol{\phi}),$$

where f_0 and g_0 are arbitrary functions, then the marginal reference prior for $\boldsymbol{\xi}$ is

$$\pi^R(\boldsymbol{\xi}) \propto f_0(\boldsymbol{\xi}).$$

Using this result, it is proved in Appendix A that the marginal reference prior of $\boldsymbol{\xi}$ is

$$\pi^R(\boldsymbol{\xi}) \propto \left|\mathbf{X}^T_{\boldsymbol{\xi}} \mathbf{X}_{\boldsymbol{\xi}}\right|^{-1/2}. \tag{5}$$

This is in general an improper distribution, depending on the choice of the experimental design.

2.2 Reference posterior distribution of $\boldsymbol{\xi}$

The reference posterior distribution of $\boldsymbol{\xi}$ is obtained by integrating out the nuisance parameter from the joint reference posterior,

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

where $\mathbf{D}$ denotes the data matrix, $\Phi = (0, \infty) \times (-\infty, \infty)^{p'}$ is the nuisance parameter space and

$$\mathrm{L}(\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2| \mathbf{D}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\boldsymbol{\alpha})^T (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\boldsymbol{\alpha}) \Big\} \tag{6}$$

is the likelihood function for the complete vector of parameters. The quadratic form in the exponent of (6) can be usefully rewritten as

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

where $\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = (\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}})^{-1} \mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{y}$ is the MLE of the nuisance parameter $\boldsymbol{\alpha}$, written as function of $\boldsymbol{\xi}$. Since $\pi^R(\boldsymbol{\alpha}|\boldsymbol{\xi}, \sigma^2, \mathbf{D}) = \mathrm{MN}\big(\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \sigma^2 (\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}})^{-1}\big)$, where MN denotes the multinormal distribution, the reference posterior distribution for $(\boldsymbol{\xi}, \sigma^2)$ is given by

$$\pi^R(\boldsymbol{\xi}, \sigma^2|\mathbf{D}) \propto \frac{\big|\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}}\big|^{-1}}{(\sigma^2)^{\frac{n-p'+2}{2}}} \exp\Big\{ -\frac{(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}))^T (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}))}{2\sigma^2} \Big\}.$$

Then, since $\pi^R(\sigma^2|\, \boldsymbol{\xi}, \mathbf{D}) = \mathrm{IG}\big((\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}))^T (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi})), n - p'\big)$, where IG denotes the inverse gamma distribution, by integrating out σ^2, it results that the posterior distribution of $\boldsymbol{\xi}$ is

$$\pi^R(\boldsymbol{\xi}|\mathbf{D}) \propto \big|\mathbf{X}_{\boldsymbol{\xi}}^T \mathbf{X}_{\boldsymbol{\xi}}\big|^{-1} \Big[(\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}))^T (\mathbf{y} - \mathbf{X}_{\boldsymbol{\xi}}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi})) \Big]^{-\frac{n-p'}{2}}. \tag{7}$$

If $k = 1$ cumbersome calculations show that (7) 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 $k = 2, 3$.

Finally, note that the marginal reference posterior distribution of $\boldsymbol{\xi}$ can be computed up to a normalizing constant. Therefore we can make posterior inference on the true optimum point by using the MCMC technique, resorting in particular to the Metropolis-Hastings algorithm.

3 Analysis of $\boldsymbol{\pi^R(\xi)}$ when $\boldsymbol{k = 2}$

The marginal reference prior for the stationary point depends on the experimental design and on its properties. When fitting second order response surface, the actual orientation of the system is generally unknown and an important property to take into account in choosing the design is rotatability. It assures that the sampling 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". The most used second order design in response surface studies is the central composite design (CCD). Given k factors, it involves the use of a 2^k factorial or fractional factorial design augmented by $2k$ axial points with coordinates $(\pm\alpha, 0, ..., 0)$, $(0, \pm\alpha, 0, ..., 0)$, ..., $(0, ..., 0, \pm\alpha)$ and n_c experimental runs at the center point. 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

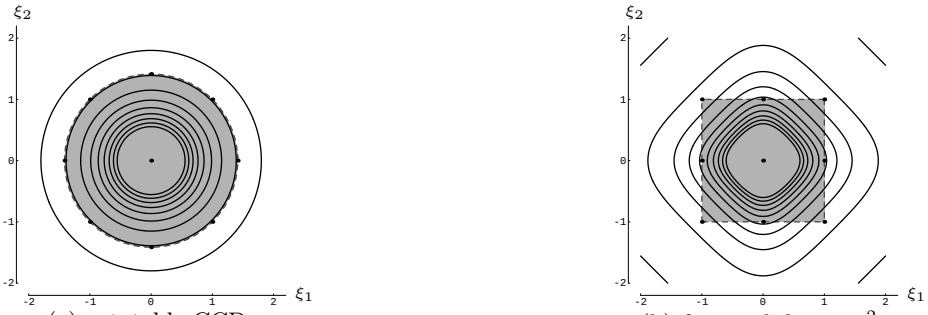
fewer experimental runs and its flexibility. In fact the CCD can be made orthogonal (the estimated effects are all non-correlated) or rotatable only choosing the location of its axial points.

When $k = 2$, the reference prior of $\boldsymbol{\xi}$ for a rotatable CCD or a factorial 3^2 design are, respectively,

$$\text{rotatable CCD:} \qquad \pi^R(\boldsymbol{\xi}) \propto \left[\left(1 + 2\xi_1^2 + 2\xi_2^2\right)\left(1 + \left(8/n_c + 3\right)\,\xi_1^2 + \left(8/n_c + 3\right)\,\xi_2^2\right)\right]^{-\frac{1}{2}}, \tag{8}$$

$$\text{factorial } 3^2\text{:} \qquad \pi^R(\boldsymbol{\xi}) \propto \left[2 + 27\left(\xi_1^2 + \xi_2^2\right) + 36\left(\xi_1^4 + \xi_2^4\right) + 288\,\xi_1^2\,\xi_2^2\right]^{-\frac{1}{2}}. \tag{9}$$

Note that (8) is written as a function of the number of center runs. They are two improper distributions, whose contours are plotted in Figure 1.



(a) rotatable CCD (b) factorial design 3^2

Figure 1: Contour plot of unnormalized $\pi^R(\boldsymbol{\xi})$, when $k = 2$

The gray areas represent the experimental regions: for the rotatable CCD it is a circle of radius $\sqrt{2}$. It is possible to see that if we use a rotatable CCD the corresponding distribution $\pi^R(\boldsymbol{\xi})$ is constant on circles around the design center. Thus, these points are considered *a priori* as "equally probable". We can say that the design rotatability is preserved from a Bayesian point of view and $\pi^R(\boldsymbol{\xi})$ is coherent with the design choice, which itself expresses the prior knowledge about the location of the stationary point. Note that the just described behavior of this prior does not depend on the number of center runs. The same situation does not hold if we adopt a 3^2 factorial plan, which is orthogonal, "in the sense that no two of the estimates for first and second order effects are correlated" (see Box and Draper, 1987), but not rotatable. However, in both cases, $\pi^R(\boldsymbol{\xi})$ presents its unique mode at the design center, $\boldsymbol{\xi} = (0, 0)$.

Moreover, let us remark that the reference prior of $\boldsymbol{\xi}$ has an infinite integral over the whole parameter space $\mathbb{R}^2$. However, it results in a proper distribution over a bounded region of operability, which can be defined when we have a real experiment to perform. Once $\mathcal{R}_O$ is fixed, the ratio

$$C(\mathcal{R}, \mathcal{R}_O) = \frac{\int_{\mathcal{R}} \pi^R(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}}{\int_{\mathcal{R}_O} \pi^R(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}}, \tag{10}$$

yields the prior probability assigned to the experimental region $\mathcal{R}$ conditional to $\mathcal{R}_O$. For simplicity, let us suppose that a rotatable CCD with n_c center runs is used and a circular region of operability with center at the design center and radius $r_o > \sqrt{2}$ is defined. Table 1 shows the ratios (10), for different values of r_o and n_c. As it can be expected, for a fixed n_c, the larger the dimension of $\mathcal{R}_O$ the smaller the reference prior probability assigned to $\mathcal{R}$ is. Once $\mathcal{R}_O$ is fixed, an increase of n_c, i.e. the experimental information in the design center, produces a decrease in the prior probability assigned to the experimental region.

Table 1: Ratios (10) for different values of n_c and r_o

	r_o				
n_c	5	10	15	20	50
1	0.478	0.368	0.324	0.299	0.239
2	0.463	0.354	0.311	0.286	0.228
4	0.451	0.342	0.300	0.275	0.219
8	0.442	0.334	0.292	0.268	0.212
16	0.436	0.328	0.287	0.263	0.208
32	0.433	0.325	0.284	0.260	0.206

Acknowledgements
The author wishes to thank Ludovico Piccinato for his support and for providing very helpful suggestions on this paper.

Appendix

A Determination of $\pi^R(\boldsymbol{\xi})$

Let us denote with $\boldsymbol{\psi} = (\boldsymbol{\xi}, \boldsymbol{\phi})$ the complete vector of parameters of M_R. Fisher's information matrix with blocks corresponding to $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ is found to be

$$\mathbf{I}_{\boldsymbol{\psi}}(\boldsymbol{\psi}) = \left[\begin{array}{c|c} \mathbf{I}_{\xi\xi}(\boldsymbol{\xi},\boldsymbol{\phi}) & \mathbf{I}_{\xi\phi}(\boldsymbol{\xi},\boldsymbol{\phi}) \\ \hline \mathbf{I}^T_{\xi\phi}(\boldsymbol{\xi},\boldsymbol{\phi}) & \mathbf{I}_{\phi\phi}(\boldsymbol{\xi},\boldsymbol{\phi}) \end{array}\right] = \frac{1}{\sigma^2}\left[\begin{array}{c|cc} \mathbf{M}\mathbf{M}^T & \mathbf{M}\mathbf{X}_{\boldsymbol{\xi}} & \mathbf{0}_{(k\times 1)} \\ \hline (\mathbf{M}\mathbf{X}_{\boldsymbol{\xi}})^T & \mathbf{X}^T_{\boldsymbol{\xi}}\mathbf{X}_{\boldsymbol{\xi}} & \mathbf{0}_{(p'\times 1)} \\ \mathbf{0}^T_{(1\times k)} & \mathbf{0}^T_{(1\times p')} & \frac{n}{2\sigma^2} \end{array}\right], \tag{11}$$

where $\mathbf{M}$ is the $k \times n$ matrix, whose generic column is the $k-$dimensional vector $2\mathbf{A}(\mathbf{x}^i - \boldsymbol{\xi})$, $\forall\, i = 1, ..., n$. That is $\mathbf{M} = \left[2\mathbf{A}(\mathbf{x}^1 - \boldsymbol{\xi}), 2\mathbf{A}(\mathbf{x}^2 - \boldsymbol{\xi}), ..., 2\mathbf{A}(\mathbf{x}^n - \boldsymbol{\xi})\right]$, where $\mathbf{x}^i = (x_{i1}, x_{i2}, ..., x_{ik})^T$ is the vector of factor levels over which the response y_i is observed.

As it will be clear in the following, it is also useful to obtain the Fisher information matrix (11) using the equivalence between models M_P and M_R. When n experimental runs have been used, the second order polynomial model can be written in matrix notation as

$$M_P: \quad \mathrm{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n),$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k, \beta_{11}, ..., \beta_{kk}, \beta_{12}, ..., \beta_{k-1,k})^T$ is the $p \times 1$ vector of regression coefficients and $\mathbf{X}$ is the design matrix. Denoting with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ the whole parameter vector of models M_P and with $\boldsymbol{g}$ the non singular transformation such that $\boldsymbol{\theta} = \boldsymbol{g}^{-1}(\boldsymbol{\psi})$, we have that

$$\mathbf{I}_{\boldsymbol{\psi}}(\boldsymbol{\psi}) = \boldsymbol{J}^T_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\mathbf{I}_{\boldsymbol{\theta}}(\boldsymbol{g}^{-1}(\boldsymbol{\psi}))\boldsymbol{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi}), \tag{12}$$

where

$$\mathbf{I}_{\boldsymbol{\theta}}(\boldsymbol{g}^{-1}(\boldsymbol{\psi})) = \frac{1}{\sigma^2}\left[\begin{array}{c|c} \mathbf{X}^T\mathbf{X} & \mathbf{0} \\ \hline \mathbf{0}^T & \frac{n}{2\sigma^2} \end{array}\right]$$

is the Fisher information matrix for the parameters of M_P and

$$\boldsymbol{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi}) = \frac{\partial \boldsymbol{g}^{-1}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \begin{bmatrix} (2\mathbf{A}\boldsymbol{\xi})^T & 1 & (\boldsymbol{\xi}^T)^2 & \boldsymbol{j}^1_{(1\times s)} & 0 \\ -2\mathbf{A} & \mathbf{0}_{(k\times 1)} & \boldsymbol{J}^1_{(k\times k)} & \boldsymbol{J}^2_{(k\times s)} & 0 \\ \mathbf{0}_{(k\times k)} & \mathbf{0}_{(k\times 1)} & \boldsymbol{I}_{(k\times k)} & \mathbf{0}_{(k\times s)} & 0 \\ \mathbf{0}_{(s\times k)} & \mathbf{0}_{(s\times 1)} & \mathbf{0}_{(s\times k)} & \boldsymbol{I}_{(s\times s)} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

is the Jacobian of the inverse transformation. In matrix (13) $\boldsymbol{I}_{(k\times k)}$ is the $k \times k$ identity matrix, $s = \frac{k(k-1)}{2}$, $\boldsymbol{j}^1 = (\xi_1\xi_2, \xi_1\xi_3, ..., \xi_{k-1}\xi_k)$, $\boldsymbol{J}^1 = \text{diag}(-2\xi_1, -2\xi_2, ..., -2\xi_k)$ and

$$\boldsymbol{J}^2 = \frac{\partial(\beta_1, ..., \beta_k)}{\partial(\alpha_{12}, \alpha_{13}, ..., \alpha_{k-1,k})} = [j^2_{h,(u,l)}], \qquad \text{where} \qquad j^2_{h,(u,l)} = \begin{cases} -\xi_l & u = h \\ -\xi_u & l = h \\ 0 & l, u \neq h \end{cases}.$$

Using the well-known properties of determinants (see, for instance, Graybill, 1969), we get

$$\left|\mathbf{S}_{\xi\xi}(\boldsymbol{\xi}, \boldsymbol{\phi})\right| = \left|\mathbf{I}_{\phi\phi}(\boldsymbol{\xi}, \boldsymbol{\phi})\right| \left|\mathbf{S}_{\boldsymbol{\psi}}(\boldsymbol{\xi}, \boldsymbol{\phi})\right| = \frac{\left|\mathbf{I}_{\phi\phi}(\boldsymbol{\xi}, \boldsymbol{\phi})\right|}{\left|\mathbf{I}_{\boldsymbol{\psi}}(\boldsymbol{\xi}, \boldsymbol{\phi})\right|} = \frac{n(\sigma^2)^{-(p'+2)} \left|\mathbf{X}^T_{\boldsymbol{\xi}} \mathbf{X}_{\boldsymbol{\xi}}\right|}{2 \left|\boldsymbol{J}^T_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\right| \left|\mathbf{I}_{\boldsymbol{\theta}}(\boldsymbol{g}^{-1}(\boldsymbol{\psi}))\right| \left|\boldsymbol{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\right|}.$$

Thus, since $\left|\boldsymbol{J}^T_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\right| = \left|\boldsymbol{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\right| = \left| -2\mathbf{A}\right|$, the following factorization holds

$$\left|\mathbf{S}_{\xi\xi}(\boldsymbol{\xi}, \boldsymbol{\phi})\right|^{-1/2} = \left|\mathbf{X}^T_{\boldsymbol{\xi}} \mathbf{X}_{\boldsymbol{\xi}}\right|^{-1/2} \left[\frac{n(\sigma^2)^{-(p'+2)}}{2 \left|\boldsymbol{J}^T_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\right| \left|\mathbf{I}_{\boldsymbol{\theta}}(\boldsymbol{g}^{-1}(\boldsymbol{\psi}))\right| \left|\boldsymbol{J}_{\boldsymbol{g}^{-1}}(\boldsymbol{\psi})\right|}\right]^{-1/2} = f_0(\boldsymbol{\xi})\, g_0(\boldsymbol{\alpha}, \sigma^2),$$

with $f_0(\boldsymbol{\xi}) = \left|\mathbf{X}^T_{\boldsymbol{\xi}} \mathbf{X}_{\boldsymbol{\xi}}\right|^{-1/2}$. Thus, from the generalization of Proposition 3 and its Corollary of Bernardo and Ramón (1998), the marginal reference prior of the quantity of interest is (5).

References

Bayarri, M.J. (1981). Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trab. Estadist.*, **32**, 18-31.

Berger, J.O. and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of the American Statistical Association*, **84**, 200-207.

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). Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, Ed. P.K. Goel e N.S. Iyengar, 323-340. Springer, Berlin.

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

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

Bernardo, J.M. and Ramón, J.M. (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician*, **47**, 101-135.

Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian theory.* Wiley, New York.

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.

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

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.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika*, **80**, 295-304.

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

Sun, D. and Ye, K. (1995). Reference prior Bayesian analysis for normal mean products. *Journal of the American Statistical Association*, **90**, 589-597.

Ye, K. and Berger, J. (1991). Noninformative priors for inferences in exponential regression models. *Biometrika*, **78**, 645-656.