ESTIMATION OF PARAMETERS AND MISSING RESPONSES IN FIRST ORDER RESPONSE SURFACE DESIGN MODEL USING EM ALGORITHM

KUNCHAM SRINIVAS, N. CH. BHATRA CHARYULU

Department of Statistics,
University College of Science, Osmania University, Hyderabad-7, India.

(Received On: 05-06-15; Revised & Accepted On: 22-06-15)

SUMMARY

Expected Maximization algorithm maximizes pastiche estimates of parameters based on the observed sample in an iterative process. This paper attempts to estimate the parameters and missing responses using Expected Maximization algorithm if the design of the experiment satisfies the first order response surface design model.

Keywords: EM algorithm, Missing responses, Response Surface Design Model.

1. INTRODUCTION

Let there be ‘v’ factors $X_i$ (i = 1, 2,...v) each at ‘s’ levels for experimentation and D denotes the design matrix with the combination of factor levels, given by

$$D = (x_{u1}, x_{u2}, ..., x_{uv})$$  \hspace{1cm} (1.1)

where $x_{ui}$ be the level of the $i^{th}$ factor in the $u^{th}$ treatment combination (i=1, 2,...v; u =1, 2,...N). Let $Y_u$ denote the response at the $u^{th}$ treatment combination. The factor-response relationship is given by

$$E(Y_u) = f(x_{u1}, x_{u2}, ..., x_{uv})$$  \hspace{1cm} (1.2)

is called the response surface. Design used for fitting the response surface model is termed as ‘response surface design’ and the model is called response surface design model.

Suppose it is required to fit a first order response surface design model expressed in the form

$$\hat{Y} = X\beta + \varepsilon$$  \hspace{1cm} (1.3)

Where $Y = (Y_1, Y_2, ..., Y_N)'$ is the vector of responses,

$X_u = (1, x_{u1}, x_{u2}, ..., x_{uv})$ is the $u^{th}$ row of X

$\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_v)'$ is the vector of parameters

$\varepsilon = (\varepsilon_1, \varepsilon_2, ... \varepsilon_N)$ is the vector of random errors and follows $N(0, \sigma^2 I)$.

The least square estimate of the parameter is given by

$$\hat{\beta} = (X'X)^{-1}X'Y$$  \hspace{1cm} (1.4)

with $\text{Var}(\hat{\beta}) = (X'X)^{-1}\sigma^2$. Then the estimated responses can be obtained from the fitted model as

$$\hat{Y} = X\hat{\beta}$$  \hspace{1cm} (1.5)

Even in some well-planned experiments and situations, the responses may not be available due to natural or manmade causes. If an observation lacks the resulting data, it is difficult to carry out the analysis as per the original plan of the experiment, it also may affect the orthogonality in case of RBD and LSD etc. Allan and Wishart (1930) initially made an attempt to estimate the missing response value in case of Randomized Block and Latin Square design using the least squares method. In case of more than one missing values, Yates (1933) developed an iterative process starting with some initial guess values. Healy and Westmacott (1956) described a more general iterative method for estimating the missing values. Later several authors made attempts on the estimation of missing values in design and analysis of experiments.

Corresponding Author: Kuncham Srinivas, Department of Statistics, University College of Science, Osmania University, Hyderabad-7, India.
2. EXPECTED MAXIMIZATION ALGORITHM

Expected Maximization algorithm maximizes pastiche estimates of parameters based on the observed sample in an iterative process. It is numerically stable and at each iteration increases the likelihood and, linearly converging.

Let \( y = [y_1, y_2, \ldots, y_n]' \) be the vector of observed sample of size ‘n’ correspondingly at design points X. Assume the sample drawn from a population is Normal with density function \( f(y) \). Assume the data may contain some unobserved or latent variable and unknown parameters. Let \( L(y) \) be the likelihood function and \( \log L(y) \) be the log of likelihood function of the sample. First estimate the values of the parameters by maximizing the likelihood function based on the known observed sample. Then evaluate the expected value of log of likelihood function. The improved version of the parameter that maximizes the expected value of Log of Likelihood function can be evaluated by repeating the above two steps of evaluations until two successive iterations will results same value or with negligible difference.

No significant work is found on the estimation of missing values using expected maximization algorithm in the experimental design directly. Therefore, this paper attempts to develop the procedure to estimate the parameters and missing responses in case of first order response surface design model using expected maximization algorithm.

3. ESTIMATION OF PARAMETERS AND MISSING OBSERVATIONS IN RESPONSE SURFACE DESIGN MODEL USING EM ALGORITHM

Let \( y = [y_1, y_2 \ldots y_n]' \) be the vector of responses correspondingly at the design matrix \( X_{nm(v+1)} \). Assume the factor-response relationship is linear model with first order response surface model in v factors, satisfying the model (1.3).

Assume the response variable \( Y \) follows \( N(X\beta, \sigma^2) \). Let us assume that at some design points the responses are missing. Then the model (1.3) can be expressed as

\[
\begin{bmatrix}
Y_i \\
Y_m
\end{bmatrix} =
\begin{bmatrix}
X_i \\
X_m
\end{bmatrix} \beta +
\begin{bmatrix}
\epsilon_i \\
\epsilon_m
\end{bmatrix}
\tag{3.1}
\]

where \( Y_j \) is the vector of (n-m) known observations, \( Y_m \) is the vector of ‘m’ missing observations, \( X_i \) is part of the design points corresponding to the known and \( X_m \) is corresponding to the missing observations design points. Let the error is also partitioned accordingly. The least square estimate of the parameters from the known observations is \( \hat{\beta} = (X'_i X_i)^{-1} X'_i Y_i \). Then the estimated missing observations can be obtained as \( \hat{Y}_m = X_m \hat{\beta} \).

Consider the problem of estimating the parameters and missing responses using expected maximization algorithm. The response variable \( Y \) follows \( N(X\beta, \sigma^2) \). If \( y = [y_1, y_2 \ldots y_n]' \) be the observed responses ( including missing responses) then the log of the likelihood function is

\[
L(y) = (2\pi \sigma^2)^{-n/2} \times \exp \left\{ \frac{-1}{2\sigma^2} \left[ \sum_{j=1}^{n-m} (y_j-x_j\beta)^2 + \sum_{j=n-m+1}^{n} (y_j-x_j\beta)^2 \right] \right\}
\tag{3.2}
\]

Log \( L(y) \) = \(-n/2 \log 2\pi - n/2 \log \sigma^2 - 1/2\sigma^2 \left[ \sum_{j=1}^{n-m} (y_j-x_j\beta)^2 + \sum_{j=n-m+1}^{n} (y_j-x_j\beta)^2 \right] \)

\tag{3.3}

In the expectation step, expected value of log of likelihood is evaluated, it results the conditional expectation of missing response observation as

\[
E[\log L(y)]= E[y_u / y, X] = x_u \hat{\beta}^{(k)} \quad \text{and} \quad E[y_u^2 / y, X] = (x_u \hat{\beta}^{(k)})' (x_u \hat{\beta}^{(k)}) + \sigma^2^{(k)}
\tag{3.4}
\]

In the implementation of the Maximization step, minimize the current residual sum of squares using least square method and estimate the parameters \( \beta \) and \( \sigma^2 \). The estimates of parameters \( \beta \) and \( \sigma^2 \) and missing responses \( (y_u; u = n-m+1, \ldots, n) \) can be expressed in recurrence terms is presented below.

\[
\frac{\partial}{\partial \beta} \log L(y) = 0 \Rightarrow \hat{\beta} = \left[ \sum_{j=1}^{n-m} x'_j x_j + \sum_{j=n-m+1}^{n} x'_j x_j \right]^{-1} \left[ \sum_{j=1}^{n-m} x'_j y_j + \sum_{j=n-m+1}^{n} x'_j y_j' \right]
\]

\[
\hat{\beta}^{(k+1)} = [XX]'^{-1} \sum_{j=1}^{n-m} x'_j y_j + \sum_{j=n-m+1}^{n} x'_j y_j^{(k)}
\tag{3.5}
\]

© 2015, IJMA. All Rights Reserved
$\frac{\partial}{\partial \sigma^2} \log L(y) = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n} \left[ \sum_{j=1}^{n-m} (y_j - x_j \hat{\beta})^2 + \sum_{j=n-m+1}^{n} (y_j - x_j \hat{\beta})^2 \right]$

$\Rightarrow \hat{\sigma}^2(k+1) = \frac{1}{n} \left[ \sum_{j=1}^{n-m} (y_j - x_j \hat{\beta}(k+1))^2 + \sum_{j=n-m+1}^{n} (y_j - x_j \hat{\beta}(k+1))^2 \right]$

$\Rightarrow \hat{\sigma}^2(k+1) = \frac{1}{n} \left[ (n-m)\hat{\sigma}_i^2(k+1) + m\hat{\sigma}_2^2(k+1) \right] \quad (3.6)$

$\frac{\partial}{\partial y_u} \log L(y) = 0 \Rightarrow \hat{y}_u = x_u \hat{\beta}$

$\Rightarrow \hat{y}_u = x_u \hat{\beta}^{(k+1)} \quad (3.7)$

The method of estimating the parameters and missing values using Least square and EM algorithm (with initial guess values as means or zeros) are illustrated through suitable examples in case of first order response surface design under with and without restrictions on the moment matrix.

**Example: 3.1:** Let us consider a response surface design conducted with three factors each with three levels with 27 design points given in the design matrix $X_{27x3}$. The vector of responses $Y$ corresponding at the design points are given below.

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \ -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & -1 & 1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 \ -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \ 0 & 16 & 2 & 1 & 0 & 2 & 16 & -1 \ -1 & 1 & -1 & 17 \ \end{bmatrix}$$

$Y = \begin{bmatrix} 159 & 395 & 149 & 25 & 255 & 251 & 184 & 363 & 378 & 260 & 454 & 98 & 422 & 270 & 237 & 362 & 146 & 417 & 150 & 103 & 455 & 172 & Y_{25} & 492 & 278 \ \end{bmatrix}$

The estimation of missing responses at the 12th and 25th design points are evaluated and presented below

**Case-(i):** The estimated values of parameters and missing observations at the design points are evaluated using Least square method and are presented below. For the above design matrix X,

$$(X'X_i)^{-1} = \begin{bmatrix} 0.040095141012 & -0.00016989466 & 0.0001698946650 & 0.002378525314 \ -0.00016989466 & 0.06379544682 & -0.008239891267 & 0.004247366632 \ 0.0001698946650 & -0.008239891267 & 0.063795446822 & 0.004247366632 \ 0.002378525314 & 0.004247366632 & 0.004247366632 & 0.059463132857 \ \end{bmatrix}$$

$$\hat{\beta} = (X'X_i)^{-1}X_i'Y_i \Rightarrow [ \begin{array}{l} 274.2528033 \ 45.23904179 \ 28.42762487 \ 18.32008155 \ \end{array} ]'$$

$$\hat{Y}_u = X_u \hat{\beta} \Rightarrow \hat{Y}_{(1,1,-1)} = 291.0642202 \text{ and } \hat{Y}_{(1,1,1)} = 275.7614$$

**Case-(ii):** Consider the initial guess values for missing observations at the 12th and 25th design points be $Y_{12} = Y_{25} = 0$. 

© 2015, IJMA. All Rights Reserved
Example 3.2: Let us consider a response surface design conducted with three factors each with three levels with 18 design points given in the matrix $X$. The vector of responses $Y$ corresponding at the design points are given below.

$$X = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \\
-1 & -1 & -1 & 0 & 0 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 1 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & -1 \\
\end{bmatrix}$$

$$Y = \begin{bmatrix}
2.83 & 3.25 & 3.56 & 2.53 & Y_5 & 3.19 & 2.23 & 2.65 & 3.06 & 2.57 & 3.08 & 3.5 & 2.42 & 2.79 & 3.03 & X_{16} & 2.85 & 3.12 \end{bmatrix}'$$

The estimation of missing responses at the 5th and 16th design points can be evaluated using...
Case-(i): The estimated values of parameters and missing responses at the design points are evaluated and presented below. For the above design matrix X, we can obtain

\[
(X'X) = \begin{bmatrix}
-1 & 3 & 2 & 0 & 1 \\
-3 & 0 & 0 & 11 & 6 \\
-5 & 1 & 6 & 13 \\
\end{bmatrix}
\]

\[
(X'X)_1 = \begin{bmatrix}
0.072523215404 & 0.010083746368 & 0.006665527260 & 0.024041474391 \\
0.010083746368 & 0.085455477696 & 0.005640061527 & -0.005298239617 \\
0.006655527260 & 0.005640061527 & 0.122372244060 & 0.054349683814 \\
0.024041474391 & -0.005298239617 & 0.054349683814 & 0.111661824189 \\
\end{bmatrix}
\]

Then \( \hat{\beta} = \begin{bmatrix} 2.8846738449 & 0.3144676124 & -0.020951489 & -0.2142636654 \end{bmatrix} \)

The resulting missing values at the design points are \( \hat{Y}(1001) = 2.671, \hat{Y}(1-111) = 2.336 \)

Case-(ii): Let us assume that the initial missing values at the 5th and 16th design points be \( Y_5 = Y_{16} = 2.91625 \) (The average of known responses).

\[
(X'X) = \begin{bmatrix}
18 & -3 & -2 & -3 \\
-3 & 13 & -1 & 0 \\
-2 & -1 & 12 & 7 \\
-3 & 0 & 7 & 15 \\
\end{bmatrix}
\]

\[
(X'X)_1 = \begin{bmatrix}
0.060139661 & 0.014322360 & 0.005771697 & 0.009334473 \\
0.014322360 & 0.081017529 & 0.010260793 & -0.001923899 \\
0.005771697 & 0.010260793 & 0.116075246 & -0.053014109 \\
0.009334473 & -0.001923899 & -0.053014109 & 0.093273479 \\
\end{bmatrix}
\]

The estimated parameters and missing observations values can be evaluated using (3.5), (3.7) at each iteration are presented in table 3.3.

<table>
<thead>
<tr>
<th>Iteration No.</th>
<th>Estimated Parameters Values</th>
<th>Estimated Missing Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta}_0 )</td>
<td>( \hat{\beta}_1 )</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>2.90196014</td>
<td>0.30640878</td>
</tr>
<tr>
<td>2</td>
<td>2.89020526</td>
<td>0.31206061</td>
</tr>
<tr>
<td>3</td>
<td>2.88643067</td>
<td>0.31371751</td>
</tr>
<tr>
<td>4</td>
<td>2.88523076</td>
<td>0.31423105</td>
</tr>
<tr>
<td>5</td>
<td>2.88485027</td>
<td>0.31439279</td>
</tr>
<tr>
<td>6</td>
<td>2.88472974</td>
<td>0.31444390</td>
</tr>
<tr>
<td>7</td>
<td>2.88469159</td>
<td>0.31446008</td>
</tr>
<tr>
<td>8</td>
<td>2.88467943</td>
<td>0.31446525</td>
</tr>
<tr>
<td>9</td>
<td>2.88467561</td>
<td>0.31446685</td>
</tr>
<tr>
<td>10</td>
<td>2.88467439</td>
<td>0.31446739</td>
</tr>
<tr>
<td>11</td>
<td>2.88467400</td>
<td>0.31446753</td>
</tr>
<tr>
<td>12</td>
<td>2.88467394</td>
<td>0.31446759</td>
</tr>
<tr>
<td>13</td>
<td>2.88467390</td>
<td>0.31446760</td>
</tr>
<tr>
<td>14</td>
<td>2.88467390</td>
<td>0.31446760</td>
</tr>
<tr>
<td>15</td>
<td>2.88467390</td>
<td>0.31446760</td>
</tr>
</tbody>
</table>

Table-3.3
Case-(iii): Let us assume that the initial missing values at the 5th and 16th design points be $Y_5 = Y_{16} = 0$. The estimated values of the parameters and missing values at each iteration are presented in table 3.4.

<table>
<thead>
<tr>
<th>Iteration No.</th>
<th>Estimated Parameters values</th>
<th>Estimated Missing Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_0$</td>
<td>$\hat{\beta}_1$</td>
</tr>
<tr>
<td>0</td>
<td>2.55689326</td>
<td>0.41762505</td>
</tr>
<tr>
<td>1</td>
<td>2.82588508</td>
<td>0.32768841</td>
</tr>
<tr>
<td>2</td>
<td>2.87462406</td>
<td>0.31870670</td>
</tr>
<tr>
<td>3</td>
<td>2.88149212</td>
<td>0.31581468</td>
</tr>
<tr>
<td>4</td>
<td>2.88366615</td>
<td>0.31484968</td>
</tr>
<tr>
<td>5</td>
<td>2.88435463</td>
<td>0.31460293</td>
</tr>
<tr>
<td>6</td>
<td>2.88457273</td>
<td>0.31451045</td>
</tr>
<tr>
<td>7</td>
<td>2.88464186</td>
<td>0.31448118</td>
</tr>
<tr>
<td>8</td>
<td>2.8846637</td>
<td>0.31447190</td>
</tr>
<tr>
<td>9</td>
<td>2.8846706</td>
<td>0.31446900</td>
</tr>
<tr>
<td>10</td>
<td>2.88467283</td>
<td>0.31446802</td>
</tr>
<tr>
<td>11</td>
<td>2.88467283</td>
<td>0.31446802</td>
</tr>
<tr>
<td>12</td>
<td>2.88467354</td>
<td>0.31446773</td>
</tr>
<tr>
<td>13</td>
<td>2.88467381</td>
<td>0.31446764</td>
</tr>
<tr>
<td>14</td>
<td>2.88467381</td>
<td>0.31446764</td>
</tr>
<tr>
<td>15</td>
<td>2.88467381</td>
<td>0.31446764</td>
</tr>
</tbody>
</table>

Table-3.4

Note:
1. The estimated values for the parameters and missing responses using the least square method and EM algorithm are resulting to same.
2. The number of iterations depend on the chosen initial values for the estimated responses.

ACKNOWLEDGEMENTS
The authors are thankful to the UGC for providing the fellowship to carry out this research work under the BSR RFSMS scheme.

REFERENCES
10. Yates F.Y. The analysis of replicated experiments when the field results are incomplete, Empire Journal of Experimental Agriculture, 1 (1933), 129-142.

Source of support: UGC, BSR RFSMS scheme, India. Conflict of interest: None Declared

[Copyright © 2015. This is an Open Access article distributed under the terms of the International Journal of Mathematical Archive (IJMA), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.]