# Design Criterion for Estimating Mean and Variance Functions

Yong B. Lim
Dept. of Statistics, Ewha Womans University, Seoul 120-750, Korea

## Abstract

In an industrial process, the proper objective is to find the optimal operating conditions with minimum process variability around the target. Vining and Myers(1990) suggest to use the separate model for the mean response and the process varian linear predictor $\tau_i = \log \sigma_i^2$ is unknown and should be estimated. Noting that the variance of $\hat{\tau}_i$ is heterogeneous, another appropriate D-optimality criterion $D_3$ based on the method of generalized least squares is proposed in this paper.

## 1. Introduction

In an industrial process, the proper objective is to find the optimal operating conditions that achieve some target value for the expected value of the response with minimum process variability. Taguchi(1986) proposes the robust parameter design methods that place a great detail of emphasis on the proper choice of levels of controllable factors in a process for variability reduction around the target. Nair(1992) and Myers et al(1992) provide overviews of the Taguchi's parameter design methods and they point out several shortcomings to the Taguchi procedure. To overcome them, the various statistical alternatives (Welch et al(1990), Vining and Myers(1990), Box and Jones(1992), ect.) have been proposed. Vining and Myers(1990) suggest to use the separate model for the mean response and the process variance. After they fit models by the method of least squares, they use the dual response approach of optimizing a primary response function while satisfying conditions on a secondary response function. Noting that the dual response approach is based on the estimated regression coefficients of those two models, how reliable optimum operating conditions by the dual response approach are depends heavily on the precision of those estimates.

Vining and Schaub(1996) propose an appropriate D-optimality criterion $D_1$ which will allow experimenters to design experiments efficiently so that the estimates of regression coefficients are precise in a sense. Vining et al(2000) propose an alternative D-optimality criterion $D_2$ by noting that all $n$ of the experimental runs are used to estimate the mean response, but $n_v$ distinct settings that are replicated are used to estimate the process variance.

In practice, the linear predictor $\tau_i = \log \sigma_i^2$ for the variance model is unknown and is estimated by $\hat{\tau}_i = \log s_i^2$, where $s_i^2$ is the observed sample variance. Noting that the variance of $\hat{\tau}_i$ is heterogeneous, another appropriate D-optimality criterion $D_3$ based on the method of generalized least squares is proposed in this paper.

For the quadratic polynomial model for the mean response and the first order for the log transformed variance over a k-dimensional cube, Vining and Schaub (1996) and Vining et al (2000) illustrate that fully replicated Central Composite Designs (CCDs) are highly efficient with respect to both $D_1$ and $D_2$ criterion. Based on $D_3$ criterion, we practically recommend CCDs with two replicates at star points and four replicates at factorial points since those differently replicated CCDs are better than fully replicated CCDs with four replication even though the size of the experiment of the former design is smaller than that of the latter design.

## 2. Reviews on design criteria for estimating mean model and variance model

It is assumed that for each design point $x_i$ in the k-dimensional controllable factors space a response variable $Y(x_i)$ can be observed. The variable $Y(x_i)$ has expected value

$$EY(x_i) = \beta' f(x_i)$$

and

$$VarY(x_i) = \sigma_i^2,$$

where $f(x_i)$ is a Kx1 vector representing the appropriate polynomial expansion of $x_i$ and $\beta$ is a Kx1 vector of unknown regression coefficients. Uncorrelated observations are taken at $x_1, \ldots, x_n$ (not necessarily distinct).

Let $z_1, \ldots, z_{n_v}$ be the distinct settings of $x_1, \ldots, x_n$ that are replicated. It is assumed that the model for the variance is

$$\tau_i = \gamma' g(z_i),$$

where $g(z_i)$ is a Sx1 vector representing the appropriate polynomial expansion of $z_i$, $\gamma$ is a Sx1 vector of unknown regression coefficients, and $\tau_i$ is a linear predictor related to the process variance $\sigma_i^2$ by

$$\sigma_i^2 = h(\tau_i).$$

Let $y$ be the $n$x1 vector of observed responses and let $\tau$ be the $n_v$x1 vector of the linear predictors. Then, the model for the mean response is

$$Ey = X\beta,$$

where $X' = [f(x_1), \dots, f(x_n)]$ and the model for the variance is

$$\tau = Z\gamma,$$

where $Z' = [g(z_1), \dots, g(z_{n_v})]$.

Consider the joint estimation of $\beta$ and $\gamma$. It is shown in Vining and Schaub (1996) that the expected information matrix, J, is

$$J = \begin{bmatrix} X' W_{11} X & 0 \\ 0 & Z' W_{22} Z \end{bmatrix},$$

where $W_{11}$ and $W_{22}$ are diagonal matrix with nonzero elements $1/\sigma_i^2$ and $(h_i'/\sigma_i^2)/2$, respectively. We take

$$\tau_i = \log\sigma_i^2,$$

so that diagonal elements of $W_{22}$ are constant and $\sigma_i^2$ is guaranteed to be positive. The expected information matrix per observation is

$$M = \frac{1}{n} J.$$

In the polynomial regression model, M represents a moments matrix in some sense. The D-optimal design criterion (Kiefer and Wolfowitz(1959), Fedorov(1972)) is to find a design in which the determinants of the expected information matrix $M$ is maximized and provides an experimenter with the smallest volume of confidence ellipsoid for the unknown regression coefficients under the normality assumption. The D-optimality criterion is known, by the celebrated Kiefer-Wolfowitz theorem (Kiefer and Wolfowitz (1960)), to control the variance of predicted values over the region of interests.

Note that $M$ depends on unknown $\sigma_i^2$. In the absence of any prior information on $\sigma_i^2$, Vining and Schaub (1996) assume that the variances are initially constant over the region of interest and propose an appropriate D-optimality based criterion $D_1$ which is adjusted for the number of unknown coefficients in the mean model and the variance model as follows;

$$D_1 = \frac{1}{n}[|X'X| \cdot |Z'Z|]^{\frac{1}{K+S}}.$$

They use $D_1$ for evaluating competing designs such as fully replicated CCDs, replicated factorial CCDs and Notz fully replicated designs.

Vining, Schaub and Modigh (2000) note that all $n$ of the experimental runs are used to estimate the mean response, but $n_v$ distinct settings are used to estimate the process variance. Thus, they propose an alternative moment matrix, $M^*$, defined by

$$M^* = \begin{bmatrix} \frac{1}{n} X'X & 0 \\ 0 & \frac{1}{n_v} Z'Z \end{bmatrix},$$

and the corresponding D-optimality criterion $D_2$ as follows;

$$D_2 = [(\frac{1}{n})^K (\frac{1}{n_v})^S |X'X|\,|Z'Z|]^{\frac{1}{(K+S)}}.$$

## 3. Proposed design criterion

In practice, the linear predictor $\tau_i = \log \sigma_i^2$ is unknown and should be estimated. We replace $\sigma_i^2$ by the observed sample variance $s_i^2$ to get an estimate $\hat{\tau}_i = \log s_i^2$. But,

$$Var(\log s_i^2) \simeq (\frac{1}{\sigma_i^2})^2 \cdot Var(s_i^2) = (\frac{1}{\sigma_i^4})^2 \cdot \frac{\sigma_i^4}{(n_i-1)^2} \cdot 2(n_i-1) = \frac{2}{(n_i-1)}.$$

Here $n_i$ is the number of replicates at $z_i$.

Since the number of replicates could be varied at design points $z_i$, the variance of $\hat{\tau}_i$ is not homogeneous. Therefore the generalized least squares estimator (g.l.s.e.) $\hat{\gamma}$ should be used for the estimation of the regression coefficient $\gamma$ for the variance model. The g.l.s.e. $\hat{\gamma}$ is given by

$$\hat{\gamma} = (Z' W_{22}^* Z)^{-1} Z' W_{22}^* \hat{\tau},$$

where $W_{22}^*$ is the diagonal matrix with elements $(n_i - 1)$ which is the weight at $z_i$.

Noting that unreplicated design points whose weights are equal to zeros do not give any information on $\log \sigma_i^2$, it can be said that all n of experimental runs are used to estimate the process variance. Thus, we propose another appropriate D-optimality criterion $D_3$ for estimating mean and variance function as follows;

$$D_3 = \frac{1}{n}[|X'X| \cdot |Z' W_{22}^* Z|]^{\frac{1}{K+S}}.$$

In summary, the proposed criterion $D_3$ is an alternative of $D_1$ in which the weights of design points for the estimation of $\log \sigma_i^2$ are considered.

**Remark 3.1.** Suppose $x_1, \ldots, x_m$ are distinct design points among n settings $x_1, \ldots, x_n$ of the k-dimensional controllable factors space.

We introduce the frequency counts $n_i$ for $1 \le i \le m$ as the number of times $x_i$ occurs among $x_1, \ldots, x_n$. Then the information matrix for the mean response model and the variance model are respectively,

$$\frac{1}{n} X'X = \sum_{i=1}^{m} \frac{n_i}{n} f(x_i) f(x_i)' \text{ and}$$

$$\frac{1}{n} Z' W_{22}^* Z = \sum_{i=1}^{m} \frac{(n_i - 1)}{n} g(x_i) g(x_i)'.$$

As the size of the experiment $n$ gets large, $1/n$ converges to zero. Thus the information matrix for the variance model is approximated by

$$\frac{1}{n} Z' W_{22}^* Z \simeq \sum_{i=1}^{m} \frac{n_i}{n} g(x_i) g(x_i)'.$$

Then, it is interesting to note that the approximated information matrix for the quadratic polynomial mean response model and the first order variance model is the same as the information matrix for the response surface model where we have quadratic polynomial in k quantitative variables, one qualitative variable u with two levels and all the 1st order interaction between u and k quantitative variables (Lim et al(1988)). That is to say, those two models are equivalent asymptotically.

## 4. Differently replicated CCD

We assume the quadratic polynomial model for the mean response and the first order model for the log transformed variance over a k-dimensional cube. The CCD is the most popular class of second-order designs used in practice. Among the modified class of CCDs which either replicate star points (axial points) or factorial points, Vining and Schaub (1996) and Vining et al (2000) show that fully replicated CCDs are highly efficient with respect to both $D_1$ and $D_2$ criterion. We consider the extended modification of class of CCDs which are allowed to have different size of replication at the center, star points and factorial points.

Let $r_C$ , $r_S$ and $r_F$ be the number of replicates at the center, star points and factorial points, respectively. Even though $D_3$ criterion takes the size of experiment $n$ into account, the experimenter cannot afford the large $n$ due to the experimental cost. We search for optimal number of replications for $0 \le r_C, r_S, r_F \le 10$. Table 1 list them.

The optimal number of replicates at factorial points $r_F^* = 10$ probably cause experimenters to be out of the experimental budget. Checking into values of $D_3$ for $0 \le r_C, r_S, r_F \le 10$, we practically recommend two replicates at star points and four replicates at factorial points based on the $D_3$- efficiency.

Only one center point is needed for $k = 2$ and none for $3 \le k \le 6$. $D_3$ - efficiencies of practically recommended designs for $2 \le k \le 6$ are .919, .939, .946, .954 and .959, respectively. On the other hand, those for the fully replicated designs with $r_S = r_F = 4$ are .918, .925, .942, .909 and .949. Thus $D_3$ - efficiencies of practically recommended designs are uniformly larger than the fully replicated CCDs even though the size of the experiment of the former design is smaller than that of the latter design.

## 5. Concluding remarks

We propose new appropriate D-optimality criterion $D_3$ for estimating the mean response and the log transformed variance function in which the weights of design points for the estimation of $\log \sigma_i^2$ are

**Table 1** Optimal number of replications in CCD for $0 \le r_C, r_S, r_F \le 10$

| | number of controllable factors $k$ | | | | |
|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 |
| $r_c^*$ | 4 | 0 | 0 | 0 | 0 |
| $r_s^*$ | 5 | 6 | 7 | 4 | 6 |
| $r_f^*$ | 10 | 10 | 10 | 10 | 10 |

considered. When we assume the quadratic polynomial model for the mean response and the first order polynomial model for the log transformed variance over a k-dimensional cube, we remark that our separate model for the mean and the variance are asymptotically equivalent to the response surface model where we have quadratic polynomial in k quantitative variables, one qualitative variable u with two levels and all the 1st order interaction between u and k quantitative variables in the sense that information matrices for those two model are same. Based on $D_3$ - efficiencies, we practically recommend CCD with two replicates at star points and four replicates at factorial points.

## References


1. Box, G.E.P. and Jones, S.(1992), "Designing Products that are Robust to the Environment", *Total Quality Management*, Vol. 3, pp. 265-282
2. Fedorov, V.V.(1972), *Theory of Optimal Experiments*, Translated by W.J. Studden and E.M. Klimko, Academic Press, New York
3. Kieffer, J. and Wolfowitz, J.(1959), "Optimum Designs in Regression Problems", *Annals of Mathematical Statistics*, Vol. 30, pp. 271-294
4. Kieffer, J. and Wolfowitz, J.(1960), "The Equivalence of Two Extreme Problems", *Canadian Journal of Statistics*, Vol. 12, pp. 363-366
5. Lim, Y.B., Studden, W.J. and Wynn, H.(1988), "A Note on Approximate D-optimal Designs for G $\times$ $2^M$", *Statistical Decision Theory and Related Topics IV*, Vol. 2, Springer-Verlag, New York, pp. 351-361
6. Nair, V.N. (EDITOR) (1992), "Taguchi's Parameter Design: A Panel Discussion", *Technometrics*, Vol. 34, pp. 127-161
7. Taguchi, G.(1986), *Introduction to Quality Engineering: Designing Quality into Products and Processes*, UniPub/ Kraus, White Plains
8. Vining, G.G. and Myers, R.H.(1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach", *Journal of Quality Technologies*, Vol. 22, pp. 38-45
9. Vining, G.G. and Schaub, D.(1996), "Experimental Designs for Estimating Both Mean and Variance Functions", *Journal of Quality Technologies*, Vol. 28, pp.135-147
10. Vining, G.G., Schaub, D. and Modigh, C.(2000), "Experimental Strategies for Estimating Mean and Variance Function", *Statistical Process Monitoring and Optimization*, Marcel Dekker, New York, pp.291-304
11. Welch, W.J., Yu, T.K., Kang, S.M. and Sacks, J.(1990), "Computer Experiments for Quality Control by Parameter Design", *Journal of Quality Technologies*, Vol. 22, pp. 15-22